Studies of Classical and Quantum Annealing

Thesis submitted for the degree of

Philosophiæ Doctor

Candidate:
Lorenzo Stella

Supervisors:
Prof. Giuseppe E. Santoro
Prof. Erio Tosatti

October, 24th 2005

Contents

List of Figures

List of Tables

Introduction

The idea of Quantum Annealing (QA) is a recent offspring of the celebrated *simulated thermal annealing* by Kirkpatrick *et al.* [1]. In simulated annealing, the problem of minimizing a certain cost (or energy) function in a large configuration space is tackled by the introduction of a fictitious temperature, which is slowly lowered in the course of a Monte Carlo or Molecular Dynamics simulation [1]. This device allows an exploration of the configuration space of the problem at hand, effectively avoiding trapping at unfavorable local minima through thermal hopping above energy barriers. It makes for a very robust and effective minimization tool, often much more effective than standard, gradient-based, minimization methods.

An elegant and fascinating alternative to such a Classical (simulated) Annealing (CA) consists in helping the system escape the local minima through *quantum mechanics*, by tunneling through the barriers rather than thermally overcoming them [2, 3]. Experimental evidence in disordered Ising ferromagnets subject to transverse magnetic fields showed that this strategy is not only feasible but presumably winning in certain cases [4]. In essence, in quantum annealing one supplements the classical energy (or cost) function, let us denote it by H_{cl}, with a suitable *time-dependent* quantum kinetic term, $H_{kin}(t)$, which is initially very large, for $t \leq 0$, then gradually reduced to zero in a given *annealing time* τ. The quantum state of the system $|\Psi(t)\rangle$, initially prepared in the fully quantum ground state $|\Psi_0\rangle$ of $H(t=0) = H_{cl} + H_{kin}(0)$, evolves according to the time-dependent Schrödinger equation

$$i\hbar \frac{d}{dt}|\Psi(t)\rangle = [H_{cl} + H_{kin}(t)]|\Psi(t)\rangle \ , \tag{1}$$

to reach a final state $|\Psi(t=\tau)\rangle$. A crucial basic question is then how the residual energy

$$\epsilon_{res}(\tau) = E_{fin}(\tau) - E_{opt} \ , \tag{2}$$

decreases for increasing annealing time τ. Here E_{opt} is the absolute minimum of H_{cl}, and $E_{fin}(\tau)$ is the average energy attained by the system after evolving for a time τ,

$$E_{fin}(\tau) = \frac{\langle \Psi(\tau) | H_{cl} | \Psi(\tau) \rangle}{\langle \Psi(\tau) | \Psi(\tau) \rangle} \, . \tag{3}$$

Generally speaking, this question has to do with the *adiabaticity* of the quantum evolution, i.e., whether the system is able, for sufficiently slow annealing (sufficiently long τ), to follow the instantaneous ground state of $H(t) = H_{cl} + H_{kin}(t)$, for a judiciously chosen $H_{kin}(t)$. The fictitious kinetic energy $H_{kin}(t)$ can be chosen quite freely, with the only requirement of being reasonably easy to implement. For this reason, this approach has also been called Quantum Adiabatic Evolution [5].

At the level of practical implementations on an ordinary (classical) computer, the task of following the time-dependent Schrödinger evolution in Eq. 1 is clearly feasible only for toy models with a sufficiently manageable Hilbert space [3, 5]. More interesting problems – like kinetically constrained systems of small Lennard-Jones clusters – may become affordable by employing a semi-classical approximation [6] or an effective Gaussian wave-packet dynamics [7].

However, the most challenging optimization problems remain beyond the grasp of a direct Schrödinger annealing dynamics. Indeed, optimization problems of practical interest usually involve astronomically large Hilbert spaces, a fact that calls for alternative stochastic approaches: Quantum Monte Carlo (QMC). These QMC techniques, in turn, are usually suitable to using *imaginary time* quantum evolution, where the $i\,\hbar\,\partial_t$ in Eq. 1 is replaced by $-\hbar\,\partial_t$.

Alternatively, a number of recent theoretical papers have applied Path-Integral Monte Carlo (PIMC) strategies to QA. A certain success has been obtained in several optimization problems, such as the folding of off-lattice polymer models [8, 9], the random Ising model ground state problem [10, 11], Lennard-Jones clusters optimization [12], and the Traveling Salesman Problem [13]. A PIMC-based QA strategy has proved to be a quite flexible and simple tool, providing often quantitatively better results than plain CA, for comparable computational cost.

Despite these practical successes, there is no general theory predicting the performance of a QA algorithm, in particular correlating it with the energy landscape of the given optimization problem, about which very little is known in many practical interesting cases [14]. The success, in turn, crucially depends on the type and effectiveness of the chosen kinetic energy H_{kin}, which in some way strongly influences

the effective energy landscape. This is a quite unpleasant situation, in view of the fact that it is *a priori* not obvious or guaranteed that a QA approach should do better than, for instance, CA. Indeed, for the interesting case of Boolean Satisfiability problems [15] – more precisely, a prototypical NP-complete problem such as 3-SAT – a recent study has shown that PIMC-QA performs definitely *worse* than simple CA [16]. A summary of the results of recent applications of PIMC-QA on different optimization problems is given in Chapter 1.

In order to gain understanding on these problems, we have moved one step back, and concentrated attention on the simplest textbook problems where the energy landscape is well under control: essentially, one-dimensional potentials, starting from a double-well potential, the simplest form of barrier. On these well controlled landscapes we have carried out a detailed and exhaustive comparison between quantum adiabatic Schrödinger evolution, both in real and in imaginary time, and its classical deterministic counterpart, i.e., Fokker-Planck evolution [17]. This work will be illustrated in Chapter 2.

On the same double well-potential, we have also studied the performance of different stochastic annealing approaches, both classical Monte Carlo annealing and PIMC-QA. The CA work is illustrated in Chapter 3, where we analyze the different annealing behaviors of three possible types of Monte Carlo moves (with Box, Gaussian, and Lorentzian distributions) in a numerical and analytical way. The PIMC-QA work is illustrated in Chapter 4, were we show the difficulties that a state-of-the-art PIMC-QA algorithm can encounter in describing tunneling even in a simple landscape, and we also investigate the role of the kinetic energy choice, by comparing the standard non-relativistic dispersion, $H_{kin} = \Gamma(t)p^2$, with a relativistic one, $H_{kin} = \Gamma(t)|p|$, which turns out to be definitely more effective.

In view of the difficulties encountered by PIMC-QA even in a simple double-well potential, we finally explored the capabilities of another well established QMC technique, the Green's Function Monte Carlo (GFMC), as a base for a QA algorithm. This time, we concentrated our attention on a very studied and challenging optimization problem, the random Ising model ground state search, for which both CA and PIMC-QA data are available [10, 11].

A more detailed summary of the results and achievements described in this Thesis, and a discussion of open issues, is contained in the final section 'Conclusions and Perspectives'.

Finally, in order to keep this Thesis as self-contained as possible, we include in the appendices a large amount of supplemental material.

Chapter 1

Previous Quantum Annealing studies on combinatorial optimization problems

In a quite general way, one could define a combinatorial optimization problem as the algorithmic task of minimizing a given *cost function* which depends on the configuration of variables assuming discrete values [18]: it is generally simple to map such problems on the problem of searching for the ground state of an appropriate Hamiltonian depending on Potts (or Ising) spin degrees of freedom [15, 19]. This is the case, for instance, of the Traveling Salesman Problem [20, 21, 22, 23], Number Partitioning [24], Boolean Satisfiability [25, 26], Vertex Covering [27], Graph Coloring [28], and many others.

Random instances of the problem – i.e., specific realizations of a given problem with random ingredients – are of particular interest, because they can be investigated by resorting to powerful techniques developed in the context of disordered statistical mechanics systems [29]. This physically oriented approach has often provided insight on the *typical-case* complexity of problem solving, in contrast with the more rigorous, but less informative, *worst-case* complexity theory, which constitutes one of the corner-stones of theoretical computer science [30]. The basic distinction between the **P** and **NP** complexity classes (that is, in essence, between problems for which a polynomial algorithm which is able to solve the worst-case instances is known, and those for which no polynomial algorithm is known) can sometimes be misleading.

Easy instances of **NP**-complete problems – the hardest of all the **NP** problems [31] –, can easily be found (see, e.g., Ref. [32]), while, sometimes, instances of **P** problems can take an exponential time, if local search techniques are used (see, e.g., Ref. [33]).

In the following, we shall briefly illustrate the results obtained by using a Quantum Annealing (QA) scheme based on Path-Integral Monte Carlo (PIMC) on three representative combinatorial optimization problems: searching for the ground state of a classical Random Ising Model (Sec. 1.1), the Traveling Salesman Problem (Sec. 1.2), a random Boolean Satisfiability problem (Sec. 1.3). The main idea behind the Path-Integral – see Sec. 4.1 for a brief introduction to the method, and Ref. [11] or App. C for a more detailed presentation – is to reduce the quantum partition function $Z = \mathrm{Tr}\, e^{-\frac{H}{T}}$ to a classical partition function involving P interacting replicas – known as the number of *Trotter's slices* – of the original system. This is possible by means of the so-called *Trotter theorem*:

$$\mathrm{Tr}\, e^{-\beta(H_{cl}+H_{kin})} = \lim_{P\to\infty} \mathrm{Tr}\, \left(e^{-\frac{\beta}{P}H_{cl}} e^{-\frac{\beta}{P}H_{kin}}\right)^P ,$$

where H_{cl} and H_{kin} are, respectively, the classical potential energy one aims at minimizing, and the kinetic term providing the quantum fluctuations. In practical applications, a possibly large but *finite* P is always considered (which leads therefore to an approximation of the actual partition function), and the resulting classical partition function is sampled in a rather straightforward way, using, for instance, a Metropolis Monte Carlo algorithm.

1.1 Ising Spin Glass

This section summarizes the results obtained in Refs. [10, 11]. Determining the ground state of an Ising Spin Glass can be an extraordinarily difficult task. To get an idea of the difficulty, it is enough to think that the number of possible configurations of a very small 32×32 square-lattice Ising model is of the order of 10^{308}, while the number of electrons in the universe is "just" of the order 10^{80}! It can be rigorously shown that, for a three-dimensional (3D) lattice case, the Random Ising model ground state determination belongs to the **NP-complete** complexity class [34], but here results on the simpler two-dimensional (2D) lattice case are reported, where the ground state energy E_{GS} can be calculated up to sufficiently large sizes (see Ref. [35], and

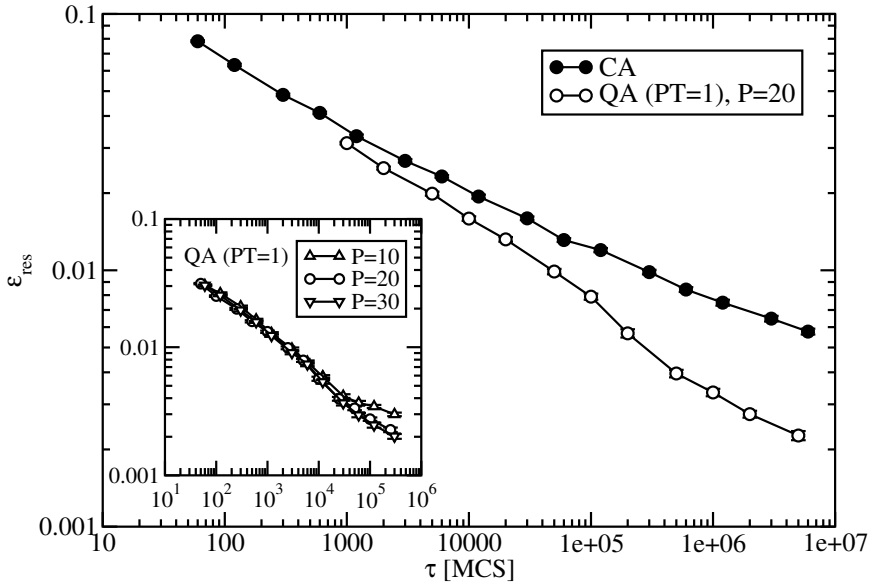

Figure 1.1: Residual energy per site for an 80×80 disordered 2D Ising model after CA and QA. QA data for the (optimal) value of $PT = 1$ are shown, with $T = 0.05$ and $P = 20$ Trotter replicas. The actual inverse annealing rate τ used in the QA has been rescaled (multiplied by P) for fair comparison with CA. Still, QA is faster than CA. (Taken from Ref. [10]).

`www.informatik.uni-koeln.de/ls_juenger/projects/sgs.html`). The Hamiltonian of the Random Ising model in a transverse field is given by:

$$H = -\sum_{\langle i,j \rangle} J_{i,j}\, \sigma_i^z\, \sigma_j^z - \Gamma \sum_i \sigma_i^x = H_{cl} + H_{kin}\ , \tag{1.1}$$

where $J_{i,j}$ are random Ising couplings between nearest-neighbor sites $\langle i, j \rangle$, σ_i^x and σ_i^z are Pauli matrices for the spin at site i, and Γ is the transverse field inducing quantum fluctuations.

For a given 2D lattice size $L \times L$, (L up to 80) and for various quenched realizations of the random couplings $J_{i,j}$, drawn from a flat distribution in the interval $(-2, 2)$, several repeated classical and quantum annealings were carried out and presented in Refs. [10, 11]. At the end of both QA and CA, the system remains generally trapped at energy $E_{final} = E_{GS} + N\epsilon_{res}$, $N = L^2$, and the efficiency of each protocol is monitored by considering the average residual energy per spin, $\epsilon_{res}(\tau)$, as

a function of the annealing time τ. The annealing parameters – the temperature T (CA) or the transverse field Γ (QA) – were decreased linearly from the initial value of $T_0 = 3$ or $\Gamma_0 = 2.5$ down to zero, in a given total number total of τ Monte Carlo steps (MCS) (i.e., complete sweeps of the whole lattice). Fig. 1.1, containing the results of a single $L = 80$ instance, shows that QA is quantitatively superior to CA in the Ising spin glass case. This numerical evidence is in agreement with the experimental observation of significantly faster frequency-dependent relaxation rates during QA of the disordered Ising ferromagnet $LiHo_{0.44}Y_{0.56}F_4$ [4].

1.2 Traveling Salesman Problem

This section summarizes the results obtained in Ref. [13] for the Traveling Salesman Problem (TSP). Given N cities and their tabulated inter-distances $d_{i,j}$, the TSP consists in finding the shortest path connecting them, visiting each city only once and returning to the starting point. An account of the vast literature about algorithms for TSP can be found *e.g.* in Ref. [36], while three classical papers analyzing physics approaches to the problem are Refs. [21, 22, 23].

As a fist step to a QA optimization, one has to chose a representation for the classical potential energy H_{cl} of a given configuration (in our case, the length of a tour), and, most crucially, a suitable source of quantum fluctuations H_{kin}. TSP was mapped [13] to a highly constrained Ising-like model – in a way similar to Refs. [3, 20] – in which each configuration of the system (a valid tour) is associated to a $N \times N$ 0/1-matrix $\hat{T}$. For every ordered sequence of cities, $\hat{T}_{i,j} = 1$ if the tour visits city i immediately after city j, and $\hat{T}_{i,j} = 0$ otherwise. For the *symmetric* TSP problem considered in Ref. [13] (a TSP with symmetric distance matrix $d_{ij} = d_{ji}$), the directed tour represented by a $\hat{T}$, and the reversed tour, represented by the transposed matrix $\hat{T}^t$, have exactly the same length. It is then convenient to introduce the symmetric matrix $\hat{U} = \hat{T} + \hat{T}^t$ as representative of *undirected* tours. The length of a tour can now be written:

$$H_{pot}(\hat{U}) = \frac{1}{2} \sum_{(ij)} d_{ij} \hat{U}_{i,j} = \sum_{\langle ij \rangle} d_{ij} \hat{U}_{i,j} \ , \tag{1.2}$$

where $\langle ij \rangle$ signifies counting each link only once. H_{kin} should be chosen in order to induce fluctuations generating the important elementary "moves" of the problem. Deciding which configurations are to become direct neighbors of a given configuration

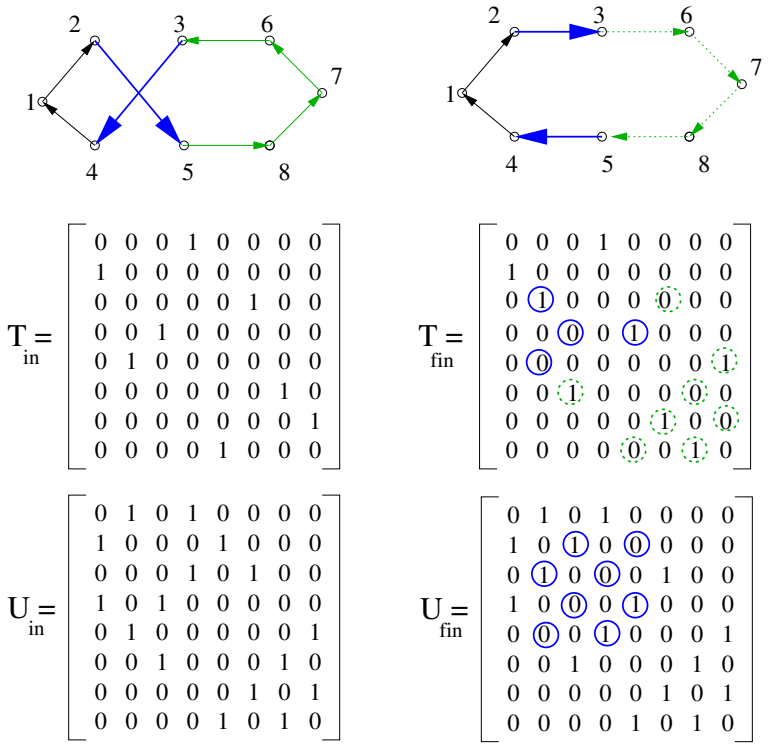

Figure 1.2: Left: Representation of an 8-city tour, with the corresponding matrix $\hat{T}_{\text{in}}$ and $\hat{U}_{\text{in}} = \hat{T}_{\text{in}} + \hat{T}_{\text{in}}^{t}$. Right: The final tour obtained when a 2-opt move is performed, with a whole section reversed (dotted line). The matrices $\hat{T}_{\text{fin}}$ and $\hat{U}_{\text{fin}}$ are shown, the circles indicating the entries that have been switched ($0 \leftrightarrow 1$) by the 2-opt move. The dotted circles in $\hat{T}_{\text{fin}}$ are entries related to the trivial reversal of a section of the tour. (Taken from Ref. [13]).

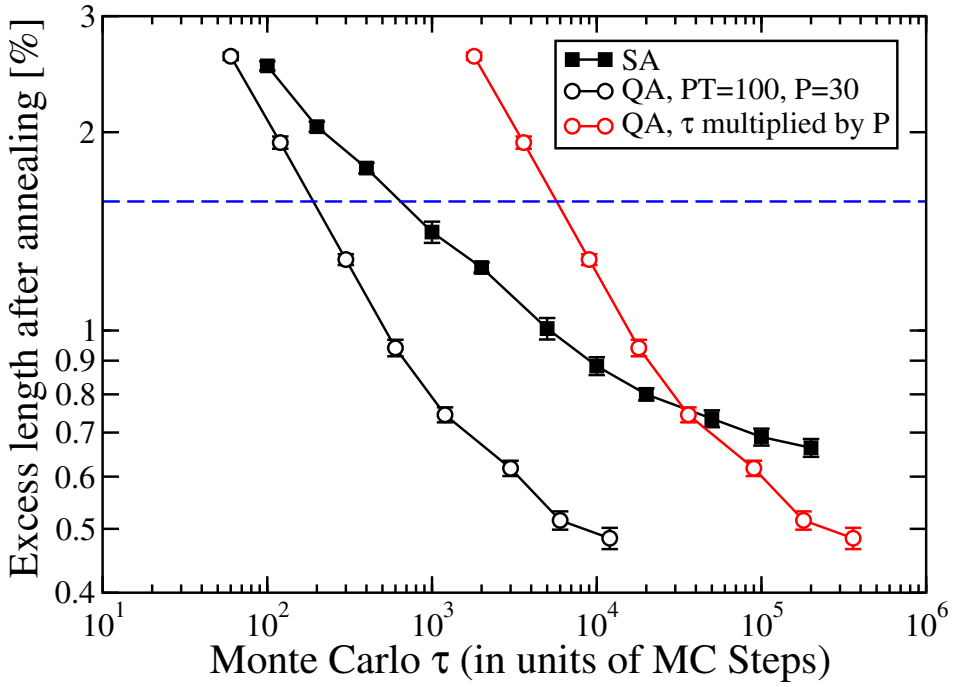

Figure 1.3: Average residual excess length found after CA and QA for a total annealing time τ (in MC steps), for the $N = 1002$ instance `pr1002` of the TSPLIB. QA is once again faster than CA. (Takef from Ref. [13]).

is indeed a crucial step, because it determines the problem's effective landscape [37]. A very important move that is often used in heuristic TSP algorithms is the so-called *2-opt move*, which consists in eliminating two links in the current tour, $(c_1 \rightarrow c_2)$ and $(c_{1'} \rightarrow c_{2'})$, and rebuilding a new tour in which the connections are exchanged, $(c_1 \rightarrow c_{1'})$ and $(c_2 \rightarrow c_{2'})$ (see Fig. 1.2). Associating a spin variable $+1$ (-1) to each entry 1 (0), the whole 2-opt move, when working with $\hat{U}$ matrices, can be represented by just four spin-flip operators:

$$S^{+}_{\langle c_{1'},c_1\rangle} S^{+}_{\langle c_{2'},c_2\rangle} S^{-}_{\langle c_2,c_1\rangle} S^{-}_{\langle c_{2'},c_{1'}\rangle} \ ,$$

where, by definition, each $S^{\pm}_{\langle i,j\rangle}$ flips an Ising spin variable (defined as $S^{z}_{\langle i,j\rangle} = (2\hat{U}_{i,j} - 1) = \pm 1$) at position (i,j) and at the symmetric position (j,i), i.e., $S^{\pm}_{\langle i,j\rangle} = S^{\pm}_{i,j} S^{\pm}_{j,i}$. However, this kinetic Hamiltonian does not allow for an obvious Trotter discretization of the Path-Integral (see discussion in Sec. 4.1), and the PIMC scheme cannot deal with it (for this purpose, Green's Function MC methods, that do not use a Trotter break-up, should be in principle more effective, see Chapter 5). One introduces then a drastic simplification to the kinetic energy term, by replacing it altogether with a standard transverse-field Ising form, arriving finally at the Hamiltonian:

$$\tilde{H}_{TSP} = \sum_{\langle ij\rangle} d_{ij} \frac{\left(S^{z}_{\langle i,j\rangle} + 1\right)}{2} - \Gamma(t) \sum_{\langle ij\rangle} [S^{+}_{\langle j,i\rangle} + H.c.] \ , \tag{1.3}$$

This simplified form of kinetic energy no longer fulfills the constraint to take a valid tour to another valid tour, but this problem is avoided by proposing exclusively 2-opt moves in the PIMC algorithm [13].

The PIMC-QA algorithm has been tested against standard CA [13] on a benchmark TSP problem, namely the printed circuit board instance `pr1002` of the TSPLIB [1]. This is a structured TSP problem with $N = 1002$ cities whose optimal tour length L_{opt} is known exactly. For CA, one can choose an optimal initial temperature T_0 by first performing several CA with various short cooling times τ and starting from sufficiently high temperatures. The point where the cooling curves for different τ's start to differ identifies an approximate "dynamical temperature" T_{dyn}. For `pr1002`, $T_{dyn} \sim 100$ was obtained. As expected [36], the optimal T_0 for CA approximately coincides with T_{dyn}. Not surprisingly, for QA the same choice $PT \sim T_{dyn}$ yields the optimal

[1] See http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95

results, together with the choice $\Gamma_0 = 300$. Fig. 1.3 shows the results obtained [13] for the average percentage best-tour excess length $\epsilon_{exc}(\tau) = (\bar{L}_{best}(\tau) - L_{opt})/L_{opt}$, both with CA (filled squares) and with QA (open circles). As a reference, the best out of 1000 runs of the Lin-Kernighan algorithm (one of the standard local-search algorithms for TSP [36]) is also plotted (dashed line in Fig. 1.3). The results show that, once again, QA anneals more efficiently, even accounting for the extra factor P in the total CPU time (rightmost open circles), reducing the error at a much steeper rate than CA.

1.3 Random Boolean Satisfiability

This final section summarized the results obtained in Ref. [16] for the 3-SAT problem. In order to state the problem, consider a set of N boolean variables $z_1, \cdots, z_N$, where $z_i = 1$ or 0 ('True' or 'False'). Denoting by ζ_i the variable z_i or its negation $\bar{z}_i$, one then considers the disjunction (logical OR) of 3 variables $C = (\zeta_i \vee \zeta_j \vee \zeta_k)$, which is called a 3- *clause*. The random 3-SAT problem consists in deciding if the conjunction (logical AND) of M different clauses $C_1 \wedge C_2 \cdots \wedge C_M$ – each clause being formed by 3 variables extracted at random among the N available, and appearing negated or not with uniform probability – can be simultaneously satisfied by a truth value assignment $\{z_i\}$. If one associates an Ising spin variable $S_i = (-1)^{z_i}$ to each Boolean variable z_i, one can assign to any clause C_a involving three variables z_i, z_j, z_k an energy E_a given by:

$$E_a = \frac{(1 + J_{a,i}S_i)\,(1 + J_{a,j}S_j)\,(1 + J_{a,k}S_k)}{8}\,, \tag{1.4}$$

where the coupling $J_{a,i}$ assumes the value -1 if the variable z_i appears negated in clause a, +1 otherwise. Evidently, $E_a = 0$ if the corresponding clause is satisfied, $E_a = 1$ otherwise.

As in the case of TSP, archives of hard structured instances exist [2]. In addition, statistical mechanics techniques can be used to determine the *phase diagram* of the Random 3-SAT problem [15, 25, 26]. The main parameter determining the hardness of a formula is the ratio $\alpha = M/N$ between the number, M, of clauses and the number, N, of variables. For $\alpha < \alpha_c \simeq 4.26$ it is typically possible to find satisfying assignments, but instances particularly hard to solve are expected to be found

[2] See http://www.intellektik.informatik.tu-darmstadt.de/SATLIB

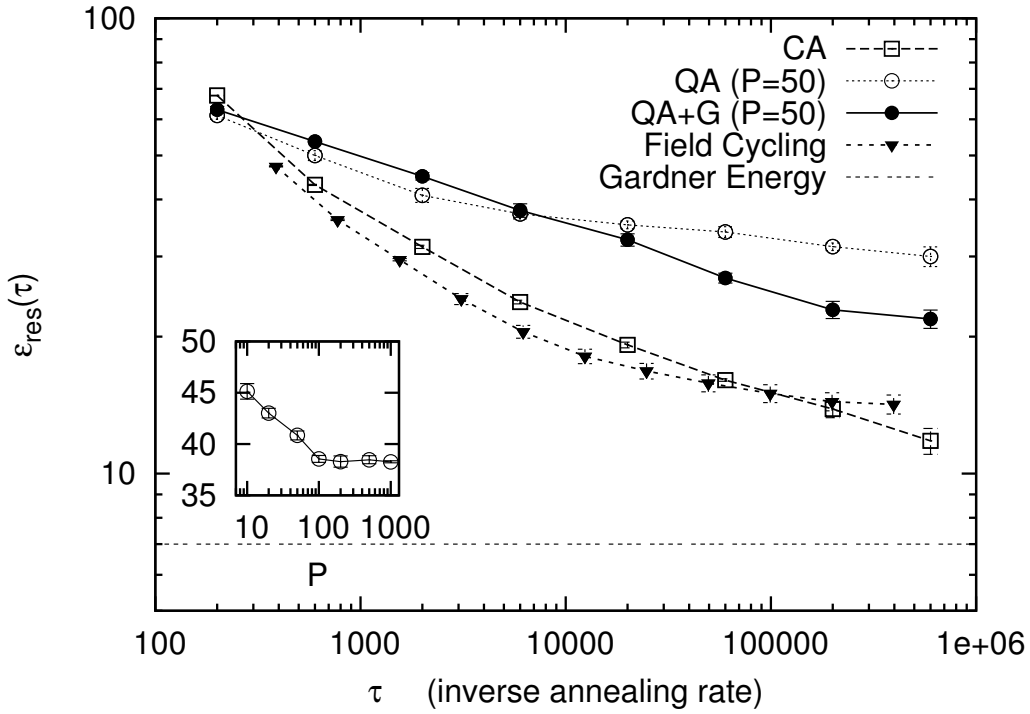

Figure 1.4: Comparison between optimal linear-schedule Classical (CA) and Quantum Annealing (QA) for a 3-SAT problem with $N = 10^4$ and $\alpha = M/N = 4.24$. CA always performs better than QA simulated with $P = 50$ Trotter replicas. The average performance of linear QA is worse than that of CA, even if an improvement in the results can be obtained by introducing global moves (G) and by increasing P (in the inset the final average energy found by QA after 2000 iterations for increasing P is plotted and compared with the average result of a CA of the same length, dashed line). The solid triangles are the data obtained by the field-cycling QA hybrid strategy discussed in Ref. [16]. (Taken from Ref. [16]).

if $\alpha > \alpha_G \simeq 4.15$ [38]. It is expected that, due to the proliferation of an exponential number of metastable states acting as dynamical traps, local search algorithms get trapped at an energy close to some finite threshold level, lower bounded by the so-called *Gardner energy* [39]. The trapping effect induced by the threshold states cannot be neglected when the instance-size is large ($N \geq 10000$) and large statistical fluctuations become sufficiently rare [38]. Smaller random formulas are, on the other hand, often easily solvable by classical simulated annealing and cannot be used as significant benchmarks.

A set of annealings was performed over a single hard 3-SAT random instance with

$N = 10^4$ and $\alpha = 4.24$ and presented in Ref. [16]. The kinetic term was chosen to be a simple transverse field inducing single-spin-flip fluctuations, like in the Ising case, since no clever sets of moves are known for 3-SAT, unlike TSP [40]. Using an efficient *ad-hoc* algorithm (that is presented in Ref. [38]), the chosen formula was tested to be actually satisfiable, as expected from theory for $\alpha < \alpha_c$. As in the case of the TSP optimization, both T_0 for CA and PT for QA have been set equal to $T_{dyn} = 0.3$. The optimal field-ramp range was found to be between $\Gamma_0 = 0.7$ and $\Gamma_f \simeq 10^{-3}$.

A comparison between the performance of the optimal CA and the optimal QA at $P = 50$, both with and without global moves (i.e., all s_i^k, $k = 1 \cdots P$ are flipped) [16], is shown in Fig. 1.4. For each point, an average was taken over 50 different realizations of the same experiment; in the case of QA, a second average was performed among the energies of the P replicas, which are in general different. It can be seen that the linear-schedule CA always performs better than the linear-schedule QA. No further improvement can be obtained for $P \geq 100$, see inset of Fig. 1.4, but $P = 50$ was chosen in order to extend as much as possible the simulation time. The asymptotic slope of the linear-schedule QA curves seems indeed to be definitely less steep than that of CA, independently of the number of replicas involved in the simulation and of the use of global moves.

Chapter 2

Annealing by Schrödinger and Fokker-Planck dynamics

As discussed in the Introduction, a simulated annealing (both thermal or quantum) takes advantage of a fictitious dynamics in order to find the global minimum of a multidimensional (discrete or continuous) energy surface (the *energy landscape* of the problem). One can introduce such a dynamics in many ways, but a natural choice is to employ a time-dependent *Fokker-Planck equation* in the classical thermal case (see below), and a time-dependent *Schrödinger's equation* in the quantum case.

In order to keep the energy landscape of the problem well under control, we concentrated our attention on simple one-dimensional potential energies, starting with the simplest case of a barrier, the double-well potential. This simple example will be highly instructive in showing the dramatic differences occurring between the two types of annealing dynamics: essentially, the height of the barrier controls the classical annealing in a standard Arrhenius activated way, while a Landau-Zener avoided-crossing tunneling transition controls the adiabaticiy loss responsible for the quantum annealing behavior.

In the rest of this chapter we shall, first, define in a more precise way the problem we want to tackle, i.e., comparing Fokker-Planck classical annealing to Schrödinger quantum annealing. Then, we shall present in detail the annealing results, both classical and quantum, for the simplest case of a potential with a barrier, a one-dimensional double-well. We shall then move to a rather amusing case of a one-dimensional potential with many minima, but no disorder, where the behavior of classical and quan-

tum annealing turns out to be remarkably different (logarithmically slow, the former, power-law fast, the latter). Finally, the crucial role played by disorder is discussed, with the example of the one-dimensional random Ising ferromagnet. Details of the calculations are contained in App. A.

2.1 Schrödinger versus Fokker-Planck annealing: Statement of the problem

Suppose we are given a potential $V(x)$, (with x a Cartesian vector of arbitrary dimension), of which we need to determine the absolute minimum (x_{opt},$E_{opt} = V(x_{opt})$). Assume generally a situation in which a steepest-descent approach, i.e., the strategy of following the gradient of V, would lead to trapping into one of the many local minima of V, and would thus not work. Classically, as an obvious generalization of a steepest-descent approach, one could imagine of performing a stochastic (Markov) dynamics in x-space according to a Langevin equation

$$\dot{x} = -\frac{1}{\eta(T)}\nabla V(x) + \xi(t) \ , \tag{2.1}$$

where the strength of the noise term ξ is controlled by the squared correlations $\overline{\xi_i(t)\xi_j(t')} = 2D(T)\delta_{ij}\delta(t-t')$, with $\bar{\xi} = 0$. Both $D(T)$ and $\eta(T)$ – with dimensions of a diffusion constant and of a friction coefficient and related, respectively, to fluctuations and dissipation in the system – are temperature dependent quantities which can be chosen, for the present optimization purpose, with a certain freedom. The only obvious constraint is in fact that the correct thermodynamical averages should be recovered from the Langevin dynamics only if $\eta(T)D(T) = k_B T$, an equality known as Einstein's relation. [1] Physically, $D(T)$ should be an increasing function of T, so as to lead to increasing random forces as T increases, with $D(T=0) = 0$, since noise is turned off at $T = 0$. Classical annealing could be in principle performed through this Langevin dynamics, by slowly decreasing the temperature $T(t)$ as a function of time, from some initially large value T_0 down to zero. Instead of working with the

[1] It is easy to see, by direct substitution in Eq. 2.2, that, at constant T, the Boltzmann distribution $e^{-V(x)/k_B T}$ is a stationary solution of the Fokker-Planck equation only if η and D satisfy the Einstein's relation, $\eta(T)D(T) = k_B T$.

Langevin equation – a stochastic differential equation – one might equivalently address the problem by studying the probability density $P(x,t)$ of finding a particle at position x at time t. The $P(x,t)$ is well known to obey a *deterministic* time-evolution equation given by the *Fokker-Planck* (FP) equation [41]:

$$\frac{\partial}{\partial t}P(x,t) \;=\; \frac{1}{\eta(T)}\,\mathrm{div}\,(P\nabla V) \;+\; D(T)\nabla^2 P\;. \tag{2.2}$$

Here, the second term in the right-hand side represents the well known *diffusion term*, proportional to the diffusion coefficient $D(T)$, whereas the first term represents the effect of the *drift* force $-\nabla V$, inversely proportional to the friction coefficient $\eta(T) = k_B T/D(T)$. Annealing can now be performed by keeping the system for a long enough equilibration time at a large temperature T_0, and then gradually decreasing T to zero as a function of time, $T(t)$, in a given *annealing time* τ. We can model this by assuming

$$T(t) = T_0 f(t/\tau)\;,$$

where $f(y)$ is some assigned monotonically decreasing function for $y \in [0,1]$, with $f(y \leq 0) = 1$ and $f(1) = 0$. In this manner the diffusion constant D in Eq. 2.2 becomes a time-dependent quantity, $D_t = D(T(t))$. The resulting time-dependent FP equation should then be solved with an initial condition given by the equilibrium Boltzmann distribution at temperature $T(t=0) = T_0$, i.e., $P(x, t=0) = e^{-V(x)/k_B T_0}$. The final average potential energy after annealing, in excess of the true minimum value, will then be simply given by:

$$\epsilon_{res}(\tau) = \int dx\, V(x)\, P(x, t=\tau) \;-\; E_{opt} \geq 0\;, \tag{2.3}$$

where E_{opt} is the actual absolute minimum of the potential V.

In a completely analogous manner, we can conceive using a Schrödinger's dynamics to perform a deterministic quantum annealing (QA) evolution of the system, by studying:

$$\xi\hbar\frac{\partial}{\partial t}\psi(x,t) = \left[-\Gamma(t)\nabla^2 + V(x)\right]\psi(x,t)\;,$$

where $\xi = i$ for a *real-time* (RT) evolution, while $\xi = -1$ for an *imaginary-time* (IT) evolution. Here $\Gamma(t) = \hbar^2/2m_t$ will be our annealing parameter, playing the role that the temperature $T(t)$ had in classical annealing. Once again, we may take $\Gamma(t)$ varying from some large value Γ_0 at $t \leq 0$ – corresponding to a small mass of the particle,

hence to large quantum fluctuations – down to $\Gamma(t = \tau) = 0$, corresponding to a particle of infinite mass, hence without quantum fluctuations. Again, we can model this with $\Gamma(t) = \Gamma_0 f(t/\tau)$, where f is a preassigned monotonically decreasing function. A convenient initial condition here will be $\psi(x, t = 0) = \psi_0(x)$, where $\psi_0(x)$ is the *ground state* of the system at $t \leq 0$, corresponding to the large value $\Gamma(t) = \Gamma_0$ and hence to large quantum fluctuations. (For such a large Γ, the ground state will be often separated by a large energy gap from all excited states.) The residual energy after annealing will be similarly given by Eq. 2.3, where now, however, the probability $P(x, t = \tau)$ should be interpreted as

$$P(x,t) = \frac{|\psi(x,t)|^2}{\int dx' \, |\psi(x',t)|^2} \, .$$

In general, the residual energy will be different for a RT or an IT Schrödinger evolution. We will discuss in some detail RT versus IT Schrödinger evolution in Sec. 2.2.1.

The basic question we pose is which annealing scheme is eventually more effective, leading to the smallest final residual energies $\epsilon_{res}(\tau)$. This might seem an ill-posed question, because the time scales involved in the classical and in the quantum evolution are different, and also because, practically, the two approaches might involve different computational costs which would imply different CPU time scales. In other words, it might seem that it only makes sense to ask how well an annealing scheme performs in a given CPU-time T_{CPU}, with all the unavoidable uncertainty associated to a CPU-time-related answer (involving, among other things, the programmer's skills, the algorithmic choices, and the computer architecture). We will show, however, that the behavior of $\epsilon_{res}(\tau)$ can be so vastly different for the different schemes, obviously in strict relation with the form of the potential, that such time scale concerns are often practically irrelevant. [2]

2.1.1 The harmonic potential: a warm-up exercise

Preliminary to any further treatment of a potential with barriers, and as a warm-up exercise which will be useful later on, we start here with the simple case of a parabolic

[2] If, for instance, $\epsilon_{res}(\tau)$ behaves in one scheme as a power-law, $1/\tau^{\alpha}$, and in another as a slowly decreasing logarithm, $1/\log\tau$, then the first scheme will be sooner or later more convenient independently of most details.

potential in one dimension, $V(x) = kx^2/2$, which has a trivial minimum in $x = 0$, with $E_{opt} = 0$, and no barriers whatsoever.

Let us consider classical FP annealing first. As detailed in the App. A.1, it is a matter of simple algebra to show that, for the harmonic potential, one can write a simple closed linear differential equation [42, 43] for the average potential energy $\epsilon_{pot}(t)$, which has the form:

$$\frac{d}{dt}\epsilon_{pot}(t) = kD_t \left[1 - \frac{2}{k_B T(t)}\epsilon_{pot}(t)\right] , \tag{2.4}$$

the initial condition being simply given by the equipartition value $\epsilon_{pot}(t = 0) = k_B T_0/2$. As with every one-dimensional linear differential equation, Eq. 2.4 can be solved by quadrature for any choice of $T(t)$ and $D_t = D(T(t))$. Assuming the annealing schedule to be parameterized by an exponent $\alpha_T > 0$, $T(t) = T_0(1 - t/\tau)^{\alpha_T}$, τ being the annealing time, and the diffusion constant $D(T)$ to behave as a power law of temperature, $D(T) = D_0(T/T_0)^{\alpha_D}$ with $\alpha_D \geq 0$, we can easily extract from the analytical solution for $\epsilon_{pot}(t)$ the large τ asymptotic behavior of the final residual energy $\epsilon_{res}(\tau) = \epsilon_{pot}(t = \tau)$. That turns out to be:

$$\epsilon_{res}(\tau) \approx \tau^{-\Omega_{CA}} \quad \text{with} \quad \Omega_{CA} = \frac{\alpha_T}{\alpha_T(\alpha_D - 1) + 1} . \tag{2.5}$$

Trivial as it is, annealing proceeds here extremely fast, with a power-law exponent Ω_{CA} that can increase without bounds (for instance if $\alpha_D = 1$) upon increasing the exponent α_T of the annealing schedule.

Consider now the Schrödinger evolution problem for this potential,

$$\begin{aligned} \xi\frac{\partial}{\partial t}\psi(x,t) &= \left[-\Gamma(t)\nabla^2 + \frac{k}{2}x^2\right]\psi(x,t) \\ \psi(x,t=0) &= \psi_0(x) , \end{aligned} \tag{2.6}$$

where $\psi_0(x) \propto \exp(-B_0 x^2/2)$ is the ground state Gaussian wavefunction corresponding to the initial value of the Laplacian coefficient $\Gamma(t = 0) = \Gamma_0$, and $\xi = i\hbar$ or $\xi = -\hbar$ for a real time (RT) or an imaginary time (IT) evolution, respectively. This problem is studied in detail in the App. A.1, where we show that a Gaussian *Ansatz* for $\psi(x,t)$, of the form $\psi(x,t) \propto \exp(-B_t x^2/2)$ with $\text{Real}(B_t) > 0$, satisfies the time-dependent Schrödinger equation as long as the inverse variance B_t of the Gaussian satisfies the following ordinary non-linear first-order differential equation:

$$-\xi\dot{B}_t \;=\; k - 2\Gamma(t)B_t^2$$

$$B_{t=0} \quad = \quad B_0 = \sqrt{\frac{k}{2\Gamma_0}} \, . \tag{2.7}$$

Contrary to the classical case, there is no simple way of recasting the annealing problem in terms of a closed linear differential equation for the average potential energy $\epsilon_{pot}(t)$. The final residual energy $\epsilon_{res}(\tau) = \epsilon_{pot}(t = \tau)$ is still expressed in terms of B_τ (or better, of its real part $\Re(B_\tau)$),

$$\epsilon_{res}(\tau) = \frac{\int dx \, V(x) |\psi(x, t = \tau)|^2}{\int dx \, |\psi(x, t = \tau)|^2} = \frac{k}{4\Re(B_\tau)} \, ,$$

but the behavior of B_t must be extracted from the study of the non-linear Eq. 2.7. The properties of the solutions of Eq. 2.7 are studied in detail in the App. A.1, where we show that:

i) $\epsilon_{res}(\tau)$ cannot decrease faster than $1/\tau$, for large τ, i.e., a power-law exponent $\epsilon_{res}(\tau) \approx \tau^{-\Omega_{QA}}$ is bound to be $\Omega_{QA} \leq 1$.

ii) Adopting a power-law annealing schedule $\Gamma(t) = \Gamma_0(1 - t/\tau)^{\alpha_\Gamma}$, the exponent Ω_{QA} for the IT case is

$$\Omega_{QA} = \frac{\alpha_\Gamma}{\alpha_\Gamma + 2} \, , \tag{2.8}$$

increasing toward the upper bound 1 as α_Γ is increased toward ∞.

iii) RT quantum annealing proceeds with exactly the same exponent Ω_{QA} as IT quantum annealing – although $\epsilon^{RT}_{res}(\tau) \geq \epsilon^{IT}_{res}(\tau)$ in general –, except that the limit $\alpha_\Gamma \to \infty$ (abrupt switch-off of the Laplacian coefficient) is singular in the RT case.

Summarizing, we have learned that, for a single parabolic valley in configuration space, both CA and QA proceed with power-laws, but CA can be much more efficient than QA, with an arbitrarily larger power-law exponent. We underline however that this is merely an academic matter at this point, steepest descent being much more efficient than both CA and QA in such a simple case. The power of QA shows up only when potentials with barriers are considered.

2.2 The simplest barrier: a double-well potential

Let us take now the classical potential to be optimized as a simple double-well potential in one-dimension

$$V_{\rm sym}(x) = V_0 \frac{(x^2 - a^2)^2}{a^4} + \delta x \ , \tag{2.9}$$

with V_0, a, and δ real constants. In absence of the linear term ($\delta = 0$), the potential has two degenerate minima located at $\pm a$, and separated by a barrier of height V_0. When a small linear term $\delta > 0$ is introduced , with $\delta a << V_0$, the two degenerate minima are split by a quantity $\Delta_V \approx 2\delta a$, the minimum at $x \approx -a$ becoming slightly favored. For reasons that will be clear in a moment, it is useful to slightly generalize the previous potential to a less symmetric situation, where the two wells possess definitely distinct *curvatures* at the minimum (i.e, their widths differ substantially). This is realized easily, with a potential of the form:

$$V_{\rm asym}(x) = \begin{cases} V_0 \frac{(x^2 - a_+^2)^2}{a_+^4} + \delta x & \text{for } x \geq 0 \\ V_0 \frac{(x^2 - a_-^2)^2}{a_-^4} + \delta x & \text{for } x < 0 \end{cases} \ , \tag{2.10}$$

with $a_+ \neq a_-$, both positive. (The discontinuity in the second derivative at the origin is of no consequence in our discussion.) To linear order in the small parameter δ, the two minima are now located at $x_\pm = \pm a_\pm - \delta a_\pm^2/(8V_0)$, the splitting between the two minima is given by $\Delta_V = \delta(a_+ + a_-)$, while the second derivative of the potential at the two minima, to lowest order in δ, is given by:

$$V''(x = x_\pm) = \frac{8V_0}{a_\pm^2} \ .$$

Obviously, $V_{\rm sym}$ is recovered if we set $a_+ = a_- = a$ in $V_{\rm asym}$.

We now present the results obtained by the annealing schemes introduced in Sec. 2.1 above. The Fokker-Planck and the Schrödinger equation (both in RT and in IT) were integrated numerically using a fourth-order Runge-Kutta method, after discretizing the x variable in a sufficiently fine real space grid. [3] (See App. A.3.) For the FP classical annealing, the results shown are obtained with a linear temperature schedule, $T(t) = T_0(1 - t/\tau)$ (i.e., $\alpha_T = 1$), and a diffusion coefficient simply

[3] Typically, we restrict the variable x in the interval $[-3, 3]$, beyond which the potential is too large and $P(x, t)$ is negligible, and use a grid of up to 1000 points, checking for convergence of the results.

proportional to $T(t)$, $D_t = D_0(1 - t/\tau)$ (i.e., $\alpha_D = 1$). Consequently, the friction coefficient is kept constant in t, $\eta_t = k_B T(t)/D_t = k_B T_0/D_0$. Similarly, for the Schrödinger quantum annealing we show results obtained with a coefficient of the Laplacian $\Gamma(t)$ vanishing linearly in a time τ, $\Gamma(t) = \Gamma_0(1 - t/\tau)$ (i.e., $\alpha_\Gamma = 1$).

Fig. 2.1 shows the results obtained for the final annealed probability distribution $P(x, t = \tau)$ at different values of τ, for both the Fokker-Planck (CA, panel (a)) and the Schrödinger imaginary-time case (IT, panel (b)), for a symmetric double well potential $V_{\rm sym}(x)$, with $V_0 = 1$ (our unit of energy), $a = a_+ = a_- = 1$ (unit of length), $\delta = 0.1$. Fig. 2.1(c) summarizes the results obtained for the residual energy $\epsilon_{res}(\tau)$. Fig. 2.2(a,b,c) shows the corresponding results for an asymmetric double well potential, Eq. 2.10, with $a_+ = 1.25$, $a_- = 0.75$, and $\delta = 0.1$.

We notice immediately that QA wins, in both cases, over CA for large enough value of τ. The RT-QA behaves as its IT counterpart for the symmetric double well, while it shows a different behavior in the asymmetric case (see below for comments). To go deeper into the details of the different evolutions, let us begin discussing the CA data (panel (a) and (c) of Figs. 2.1 and 2.2), which show similar behaviors for both choices of the potential. Starting from an initially broad Boltzmann distribution at a high $T = T_0 = V_0$, $P(x, t = 0)$ (solid lines), the system quickly sharpens the distribution $P(x, t)$ into two well-defined and quite narrow peaks located around the two minima $x_\pm$ of the potential. This agrees very well with expectations based on the CA in a harmonic potential, which showed that the width of the Gaussian should decrease linearly in τ ($\Omega_{CA} = 1$ for $\alpha_T = \alpha_D = 1$), as is indeed found in our double well case too. If we denote by $p_\pm$ the integral of each of the two narrow peaks, with $p_- + p_+ = 1$, it is clear that the problem has effectively been reduced to a *discrete* two-level system problem. The time evolution $p_\pm$ therefore obeys a discrete Master equation which involves the thermal promotion of particles over the barrier V_0, of the form:

$$\gamma^{-1}\frac{dp_+}{dt} = [1 - p_+(t)]\, e^{-\frac{\Delta_V + B}{k_B T(t)}} - p_+(t) e^{-\frac{B}{k_B T(t)}} \ , \tag{2.11}$$

where γ is an attempt frequency, while $B = V_0 - V(x_+)$ and $B + \Delta_V = V_0 - V(x_-)$ are the potential barriers seen from a particle in the metastable minimum, x_+, and in the true minimum, x_-, respectively. Eq. 2.11 was studied by Huse and Fisher in Ref. [44], where they showed that the asymptotic value of the residual energy $\epsilon_{res}(\tau) = \Delta p_+(\tau)$ is given by:

$$\epsilon_{res}(\tau) \sim \text{const} \ \ (\gamma\tau)^{-\frac{\Delta_V}{B}} (\ln \tilde{\gamma}\tau)^{2\frac{\Delta_V}{B}} \ , \tag{2.12}$$

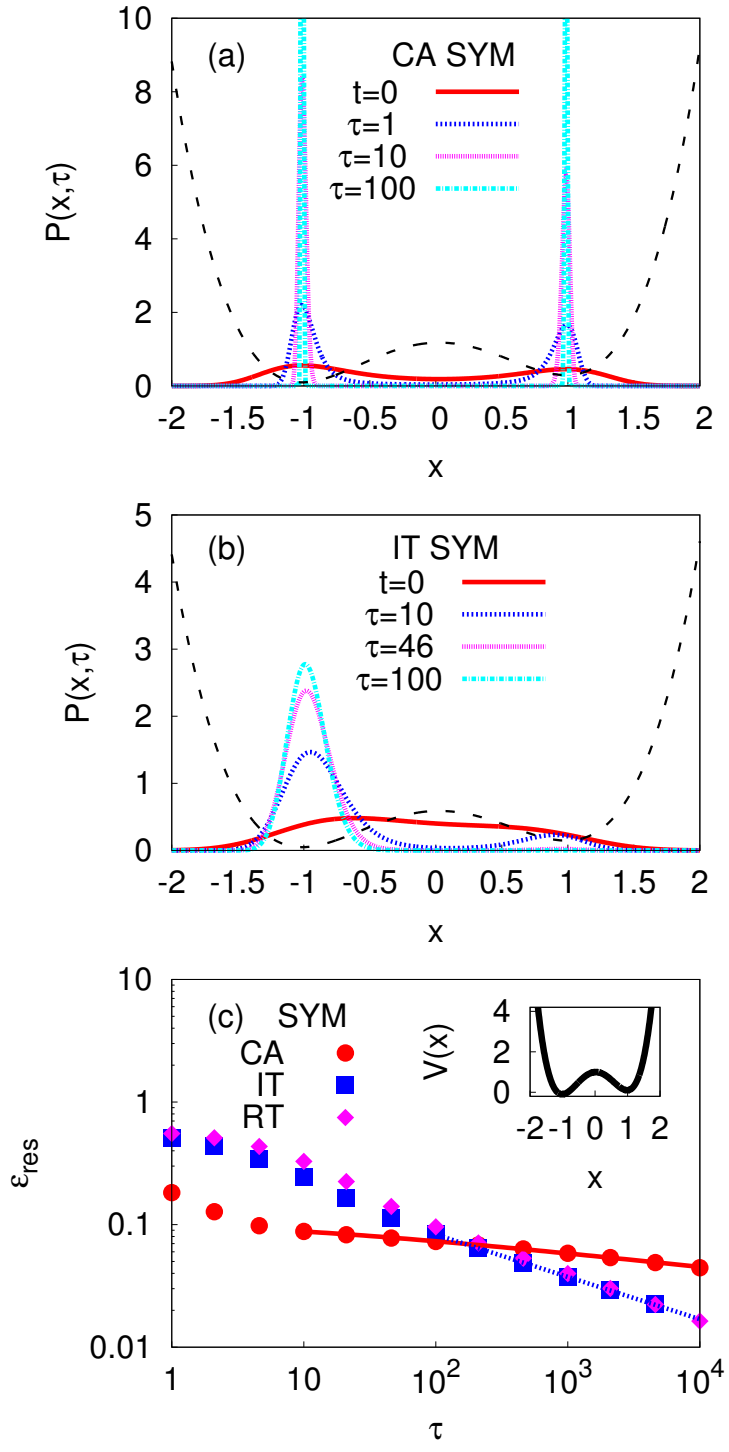

Figure 2.1: (a,b): The annealed final probability distribution $P(x, t = \tau)$ at different values of the annealing time τ, for both the Fokker-Planck classical annealing (CA, panel (a)), and the Imaginary Time Schrödinger quantum annealing (IT-QA, panel (b)). (c) Final residual energy $\epsilon_{res}(\tau)$ versus annealing time τ for quantum annealing in Real Time (RT) and Imaginary Time (IT) compared to the Fokker-Planck classical annealing (CA). The solid line in (c) is a fit of the CA data. The double well potential (dashed line in (a,b), inset of (c)) is here given by Eq. 2.10 with $a_+ = a_- = a = 1$.

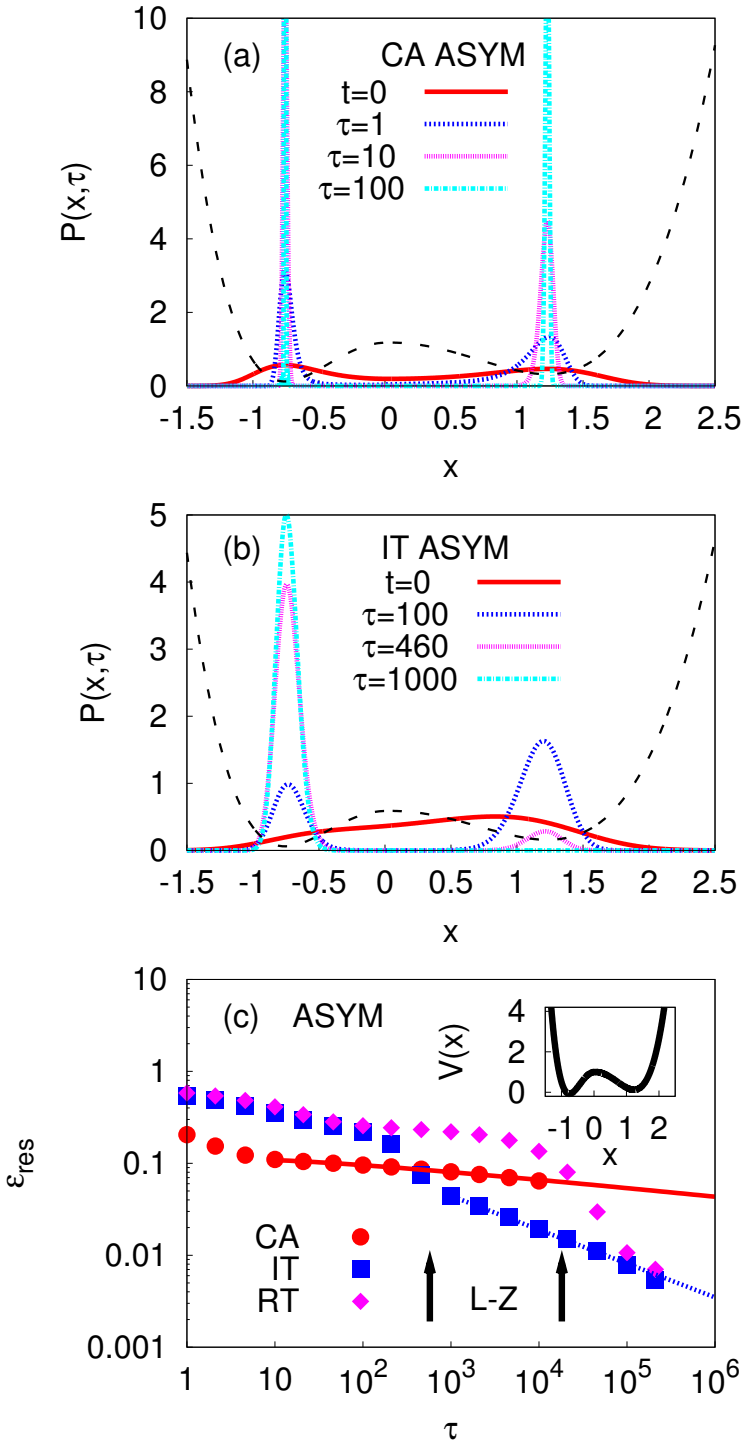

Figure 2.2: Same as Fig. 2.1, for the asymmetric potential in Eq. 2.10 with $a_+ = 1.25, a_- = 0.75$ (dashed line in (a,b), inset of (c)). Notice the different behavior of RT and IT, in the present case.

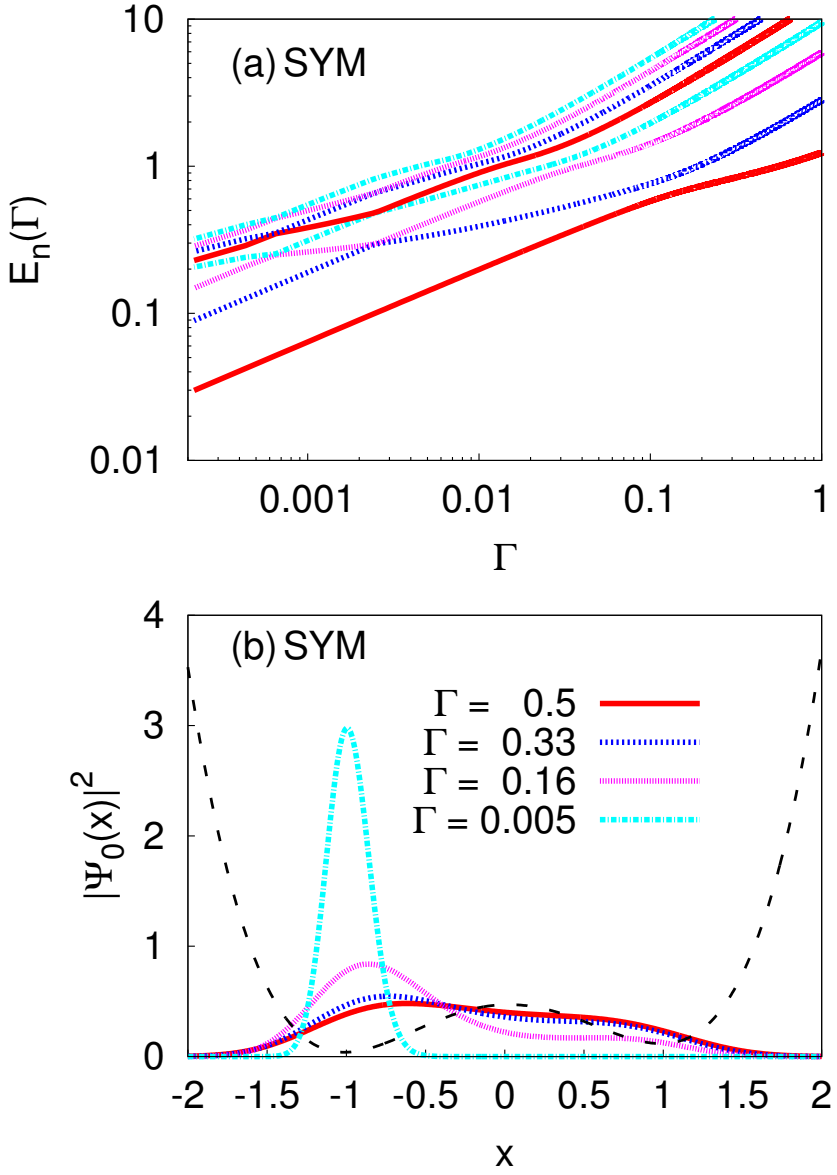

Figure 2.3: Instantaneous eigenvalues (a) and ground state wavefunctions (b) of the Schrödinger problem $H\psi = E\psi$ for different values of Γ, for the symmetric potential in Eq. 2.10 with $a_+ = a_- = a = 1$.

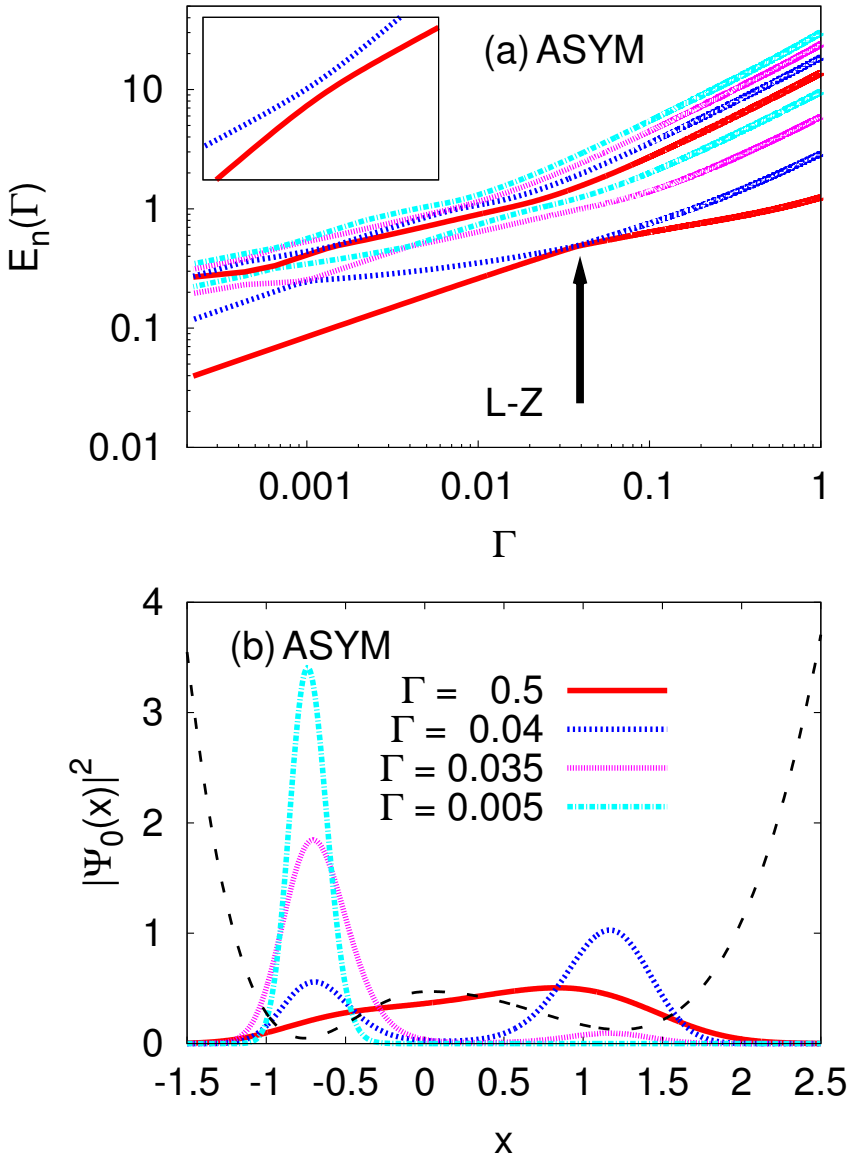

Figure 2.4: Same as Fig. 2.3, for the asymmetric potential in Eq. 2.10 with $a_+ = 1.25, a_- = 0.75$. Notice the clear Landau-Zener avoided crossing in (a), indicated by the arrow and magnified in the inset.

where $\tilde{\gamma}$ is a constant. So, apart from logarithmic corrections, the leading power-law behavior is of the form $\epsilon_{res} \sim \tau^{-\Delta_V/B}$, where the exponent is controlled by the ratio Δ_V/B between the energy splitting of the two minima Δ_V and the barrier B. As shown in Figs. 2.1(c) and 2.2(c) (solid lines through solid circles), the asymptotic behavior anticipated by Eq. 2.12 fits nicely our CA residual energy data (solid circles), as long as the logarithmic corrections are accounted for in the fitting procedure. [4] Obviously, we can make the exponent as small as we wish by reducing the linear term coefficient δ, and hence the ratio Δ_V/B, leading to an exceedingly slow classical annealing.

The behavior of the QA evolution is remarkably different. Starting from Fig. 2.1, we notice that IT and RT evolutions give very similar residual energies, definitely faster decaying than the CA data, while the corresponding final wavefunction only slowly narrows around the minimum of the potential. Notice also the asymptotic behavior of the residual energy, $\epsilon_{res}(\tau) \propto \tau^{-1/3}$, indicated by the dashed line in Fig. 2.1(c): this rather strange exponent is simply the appropriate one for the Schrödinger annealing with a linear schedule within an harmonic potential (the lower minimum valley, see Sec. 2.1.1). The asymmetric potential results, shown in Fig. 2.2, are even more instructive. The initial wavefunction squared $|\psi(x, t = 0)|^2$ corresponds to a quite small mass (a large $\Gamma_0 = 0.5$), and is broad and delocalized over both minima (solid line). As we start annealing, and if the annealing time τ is relatively short – that is, if $\tau < \tau_c$, with a characteristic time τ_c which depends on which kind of annealing, RT or IT, we perform – the final wavefunction becomes mostly concentrated on the *wrong minimum*, roughly corresponding to the ground state with a still relatively large $\Gamma_1 < \Gamma_0$ (see Fig. 2.4). The larger width of the wrong valley is crucial, giving a smaller quantum kinetic energy contribution, so that tunneling to the other (deeper) minimum does not yet occur. By increasing τ, there is a crossover: the system finally recognizes the presence of the other minimum, and effectively tunnels into it, with a residual energy that, once again, decays asymptotically as $\epsilon_{res}(\tau) \propto \tau^{-1/3}$ (dashed line in Fig. 2.2(c)). There is a characteristic annealing time τ_c – different in the two Schrödinger cases, RT and IT – above which tunneling occurs, and this shows up as the clear crossover in the residual energy behavior of both IT and RT, shown in

[4] This is generally speaking a difficult fit: The estimated value of Δ/B is between 0.14 and 0.22, depending on the inclusion of sub-leading corrections, to be compared with the theoretical value of 0.22. Neglecting the log-corrections leads to a wrong estimate of the power-law exponent.

Fig. 2.2(c).

These findings can be quite easily rationalized by looking at the *instantaneous* (adiabatic) eigenvalues and eigenstates of the associated time-independent Schrödinger problem, which we show in Fig. 2.4(a,b). Looking at the instantaneous eigenvalues shown in Fig. 2.4(a) we note a clear avoided-crossing occurring at $\Gamma = \Gamma_{LZ} \approx 0.038$, corresponding to a resonance condition between the states in the two different valleys of the potential. For $\Gamma > \Gamma_{LZ}$ the ground state wavefunction is predominantly concentrated in the wider but metastable valley, while for $\Gamma < \Gamma_{LZ}$ it is mostly concentrated on the deeper and narrower global minimum valley. In the full time-dependent RT evolution, transfer to the lower valley is a Landau-Zener problem [45, 46]: the characteristic time τ_c for the tunneling event is given by $\tau_{LZ} = \hbar\alpha\Gamma_0/2\pi\Delta^2$, where α is the relative slope of of the two crossing branches as a function of Γ, 2Δ is the gap at the avoided-crossing point, and Γ_0 is the initial value of the annealing parameter. (For the case shown in Fig. 2.4, we have $2\Delta = 0.0062$, $\alpha = 2.3$, hence $\tau_{LZ} \approx 18980$, see rightmost arrow in Fig. 2.2(c).) The Landau-Zener probability of jumping, during the evolution, from the ground state onto the "wrong" (excited) state upon fast approaching of the avoided level crossing is $P_{ex} = e^{-\tau/\tau_{LZ}}$, so that adiabaticity applies only if the annealing is slow enough, $\tau > \tau_{LZ}$. The IT characteristic time is smaller, in the present case, than the RT one. We will comment further on this point in Sec. 2.2.1. In a nutshell, the reason for this is the following. After the system has jumped into the excited state, which occurs with a probability $P_{ex} = e^{-\tau/\tau_{LZ}}$, the residual IT evolution will filter out the excited state; this relaxation toward the ground state is controlled by the annealing rate as well as by the average gap seen during the residual evolution. Numerically, the characteristic time τ_c seen during the IT evolution is of the order of $\hbar/(2\Delta)$, see leftmost arrow in Fig. 2.2(c), rather than being proportional to $1/\Delta^2$ as τ_{LZ} would imply.

Obviously, instantaneous eigenvalues/eigenvectors can be studied for the Fokker-Planck equation as well; their properties, however, are remarkably different from the Landau-Zener scenario just described for the Schrödinger case. Fig. 2.5(c) shows the first four low-lying eigenvalues of the FP equation as a function of T (for a symmetric choice of the potential), while Fig. 2.5(a,b) show the corresponding eigenstates for two value of the temperature, $T/V_0 = 1$ and $T/V_0 = 0.1$. (The asymmetric potential cases are virtually identical, and are not shown). The lowest eigenvalue of the FP operator is identically 0 and the corresponding eigenvector [41] is the Boltzmann

distribution $e^{-V(x)/k_BT}$, with roughly symmetric maxima on the two valleys. The first excited state correspond to distribution peaked on the two valley but with a node at the origin, and is separated from the ground state by an exponential small Arrhenius-like gap e^{-B/k_BT}. Higher excited states are separated by a very large gap, so that, effectively, only the two lowest lying states dominate the dynamics at small temperature. The reduction of a continuum double-well FP classical dynamics onto a discrete effectively quantum two-level system, previously noticed, is quite evident from this form of the spectrum. On the contrary, the true quantum case never allowed for a discrete two-level system description whatsoever, except perhaps for large Γ. For small enough $\Gamma < \Gamma_{LZ}$, in particular, the tower of oscillator states within the valley at x_- is always very close in energy to the actual ground state, and the quantum annealing evolution reduces effectively to a particle in a single harmonic well. This explains the rather large width of the final distributions $P(x,\tau)$ observed in the quantum case.

Summarizing, we have found that QA and CA proceed in a remarkably different way. CA is sensitive to the height of the barrier, more precisely to the ratio Δ_V/B between the energy offset Δ_V of the two minima, and the barrier height B. On the contrary, QA crucially depends on the tunneling probability between the two valleys, which is reflected in a Landau-Zener (avoided crossing) gap: a wide tunneling barrier is obviously bad for QA. Finally, we noticed that RT and IT proceed with somewhat different characteristic times: we discuss this issue a bit more in the following section.

2.2.1 Real- versus imaginary-time Schrödinger evolution

A Schrödinger dynamics in imaginary-time (IT) is clearly much more convenient than that in real-time (RT) for simulations on current classical computers, but it makes a difference in the final results? The answer to this question is, we believe: no, it does not make a difference, in essence, although IT does give quantitatively better results.

To qualify this statement, let us denote by $|\Psi^{(\xi)}(t)\rangle$ the solution of the Schrödinger equation

$$\begin{aligned} \xi\frac{d}{dt}|\Psi^{(\xi)}(t)\rangle &= [H_{cl}+H_{kin}(t)]|\Psi^{(\xi)}(t)\rangle \\ |\Psi^{(\xi)}(t_0)\rangle &= |\Psi_0\rangle \end{aligned}$$

where we assume that $|\Psi_0\rangle$ is the ground state of the initial Hamiltonian at time t_0, $H_{cl}+H_{kin}(t_0)$, while $\xi = i\hbar$ for RT or $\xi = -1\hbar$ for IT. By definition, the final

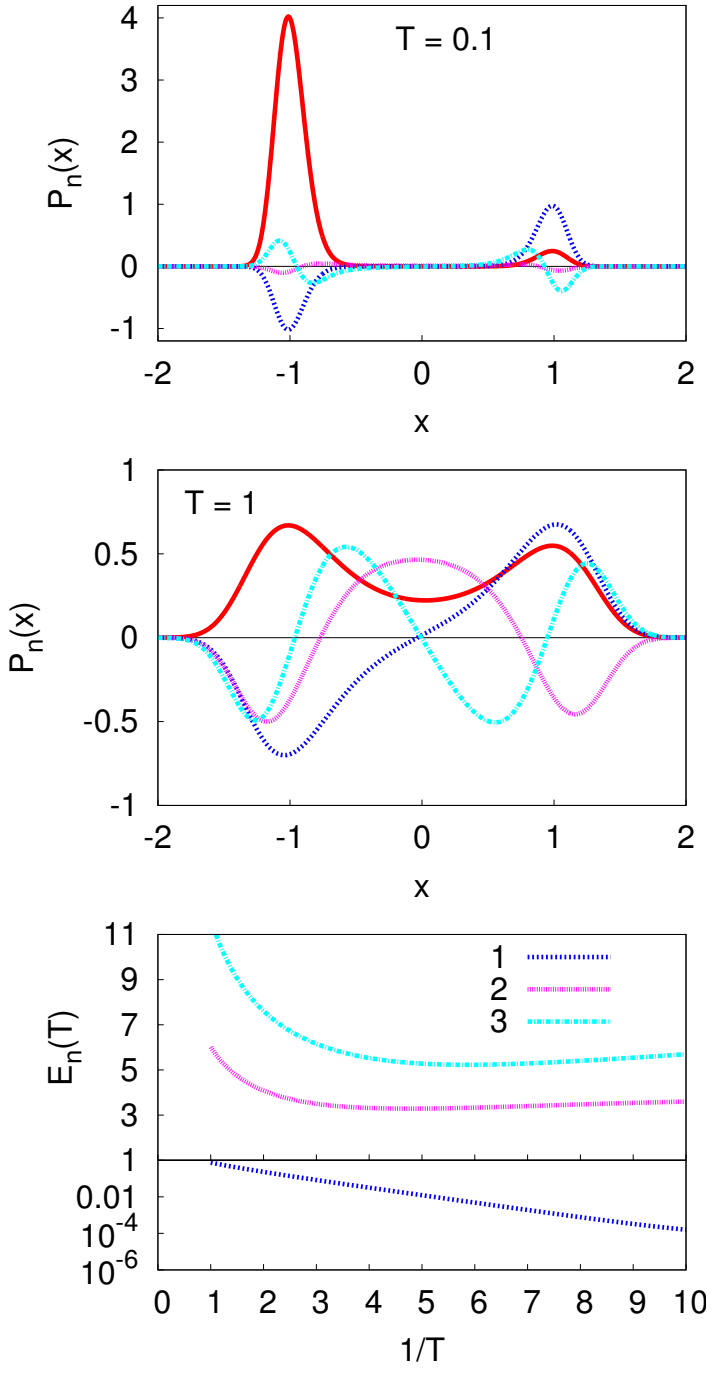

Figure 2.5: Instantaneous Fokker-Planck eigenvalues (panel (c)) as a function of temperature T, and the corresponding eigenstates for two values of T (panels (a) and (b)). The potential is here the symmetric one, $V_{\rm sym}$ in Eq. 2.9 with $V_0 = 1$, $a = 1$, $\delta = 0.1$. Similar results (not shown) are obtained for the asymmetric double well potential $V_{\rm asym}$.

residual energy after annealing up to time $t_f = t_0 + \tau$, where the kinetic energy is finally turned off, is given by:

$$\epsilon_{res}^{(\xi)}(\tau) = \frac{\langle \Psi^{(\xi)}(t_0+\tau)|H_{cl}|\Psi^{(\xi)}(t_0+\tau)\rangle}{\langle \Psi^{(\xi)}(t_0+\tau)|\Psi^{(\xi)}(t_0+\tau)\rangle} - E_{opt} \ .$$

We conjecture that the residual energies for the two alternative way of doing a Schrödinger evolution verify the following: i) the IT residual energy is *not larger* then the RT one, that is

$$\epsilon_{res}^{(IT)}(\tau) \leq \epsilon_{res}^{(RT)}(\tau) \ ,$$

and ii) in many problems, the leading asymptotic behavior, for $\tau \to \infty$, might be identical for $\epsilon_{res}^{(IT)}(\tau)$ and $\epsilon_{res}^{(RT)}(\tau)$.

Expectation (i) seems very reasonable: it is simply inspired by the time-independent case, where it is well known that the IT Schrödinger dynamics tends to "filter the ground state" out of the initial trial wave function, as long as the gap between the GS and the first excited state is non-zero. However, we have here a time-dependent situation, and the result is *a priori* not guaranteed. We do not have a proof of this statement, but we have verified it in all the cases where an explicit integration of the Schrödinger equation has been possible (see, for instance, the results of the previous section). (Needless to say, we have no proof of (ii) either, but, again, it never failed in all our tests.)

The simplest time-dependent problem where one can test our conjectures, is the discrete two-level system (TLS) problem. Here, in terms of Pauli matrices, $H_{cl} = \Delta\sigma^z$, while $H_{kin}(t) = -\Gamma(t)\sigma^x$, with $\Gamma(t) = -vt$. The full $H(t)$ is therefore

$$H(t) = \Delta\sigma^z - \Gamma(t)\sigma^x \ . \tag{2.13}$$

The annealing interpretation is very simple: the classical optimal state is $|\downarrow\rangle$, with energy $E_{opt} = -\Delta$, separated from the excited state $|\uparrow\rangle$ by a gap 2Δ. The kinetic term induces transitions between the two classical states. Starting from the ground state of $H(t_0)$ at time $t_0 = -\tau$ we let the system evolve up to to time $t_f = t_0 + \tau = 0$, when the Hamiltonian is entirely classical, $H(t_f = 0) = H_{cl} = \Delta\sigma^z$. The probability of missing the instantaneous final ground state $|\downarrow\rangle$, ending up with the excited state $|\uparrow\rangle$, is: $P_{ex}(0) = |\langle\uparrow|\Psi^{(\xi)}(0)\rangle|^2/\langle\Psi^{(\xi)}(0)|\Psi^{(\xi)}(0)\rangle$. In principle, P_{ex} depends, for given Δ, both on the initial $\Gamma(t_0) = \Gamma_0 = v\tau$ and on the annealing time τ. The really important parameter, however, turns out to be the ratio v between these two quantities,

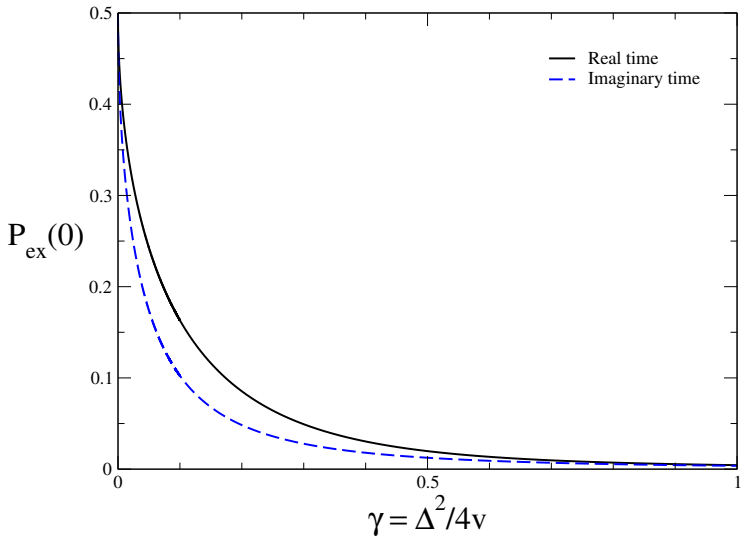

Figure 2.6: The probability $P_{ex}(0)$ of ending up into the excited state, given by Eq. 2.14, for the discrete two-level system problem in Eq. 2.13, for both imaginary-time (IT, dashed line) and real-time (RT, solid line) Schrödinger annealing. The large-γ behavior of $P_{ex}(0)$ is, in both cases, given by $P_{ex}(0) \approx 1/(256\gamma^2)$.

which determines the "velocity of annealing": Taking $\tau \to \infty$ (i.e., $t_0 \to -\infty$), and $\Gamma_0 \to \infty$ with $\Gamma(t) = -vt$ for every t, the problem can be solved analytically (in terms of parabolic cylinder functions, see for instance Ref. [3] for the RT case) for both RT and IT. The probability $P_{ex}(0)$ of ending into the excited state can be expressed in terms of the variable $\gamma = \Delta^2/4v$. The explicit expressions, in terms of Gamma functions, are:

$$\begin{aligned} P_{ex}(0) &= \frac{|R+1|^2}{2(1+|R|^2)} \\ R &= e^{i\phi_0}\frac{1}{\sqrt{\gamma}}\frac{\mathbf{\Gamma}(1+z)}{\mathbf{\Gamma}(1/2+z)}\,, \end{aligned} \tag{2.14}$$

where $\phi_0 = 3\pi/4$ and $z = i\gamma$ for RT, while $\phi_0 = \pi$ and $z = \gamma$ for IT. A plot of P_{ex} for both RT and IT is shown in Fig. 2.6 as a function of $\gamma = \Delta^2/4v$. Note that: i) the IT-result for P_{ex} (dashed line) is always below the RT-result, ii) the difference between the two curves is only quantitative: one can verify analytically that the leading

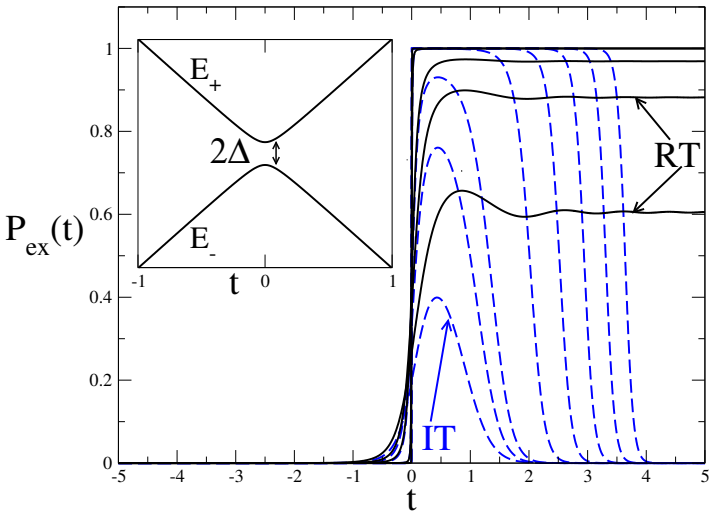

Figure 2.7: Comparison between the RT (solid lines) and the IT (dashed lines) evolution of a Landau-Zener problem, Eq. 2.15, for several values of the tunneling gap 2Δ (the values of Δ shown are $\Delta = 0.4, 0.2, 0.1, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}$, while $v = 1$). The inset shows the two instantaneous eigenvalues of the problem, $E_{\pm}(t)$, as a function of t.

behavior for large γ is the same in both cases, i.e., $P_{ex} \approx 1/(256\gamma^2)$. Similar results are obtained by direct numerical integration of the Schrödinger equation for finite Γ_0 and τ, and with other forms of $\Gamma(t)$.

With the same toy model, we can illustrate another point raised in the previous section: what happens to the IT evolution *after* a Landau-Zener avoided crossing gap is encountered. The Hamiltonian we consider is essentially that in Eq. 2.13, simply rotated in spin space,

$$H(t) = -vt\,\sigma^z - \Delta\sigma^x\ . \tag{2.15}$$

In absence of the tunneling amplitude Δ, the two energy levels would cross at $t = 0$, while for $\Delta > 0$ the two instantaneous eigenvalues are simply $E_\pm(t) = \pm\sqrt{(vt)^2 + \Delta^2}$ (see inset in Fig. 2.7). Starting with the system in the ground state at $t = -\infty$, we can monitor the probability of getting onto the excited states at any time t, which we plot in Fig. 2.7 for both the RT and the IT evolution and for several values of Δ (taking $v = 1$). The RT data provide an illustration of the well-known Landau-Zener result: after a (relatively short) tunneling time, and possibly a few oscillations, the probability of of getting onto the excited state saturates to a value given by $P_{ex}(t = \infty) = e^{-\pi\Delta^2/\hbar v}$. As for the IT data, the initial (tunneling) part and the subsequent plateau of the curves are similar to the RT case: the plateau value attained, call it P^*_{ex}, is indeed very close to the RT saturation value (in fact, asymptotically the same for $\Delta \to 0$); after that, the IT evolutions starts to filter out the ground state component – initially present in the state with a small amplitude $1 - P^*_{ex}$ – through the usual mechanism of suppression of excited states, leading to a $P_{ex}(t)$ which is nicely fit by the curve

$$P_{ex}(t) = \frac{P^*_{ex} e^{-2\int_0^t dt'[E_+(t')-E_-(t')]}}{(1-P^*_{ex}) + P^*_{ex} e^{-2\int_0^t dt'[E_+(t')-E_-(t')]}}\ ,$$

which asymptotically goes to zero as $t \to \infty$. This rather trivial effect of filtering, if on one hand explains the discrepancy between the IT and the RT evolution observed in the asymmetric double well case of the previous section, is, on the other hand, of no harm at all: on the contrary, it provides a quantitative improvement of IT over RT.

In summary, the essential equivalence of IT and RT Schrödinger annealing (with, moreover, a quantitative improvement of IT over RT) justifies practical implementations of quantum annealing based on imaginary-time Quantum Monte Carlo schemes.

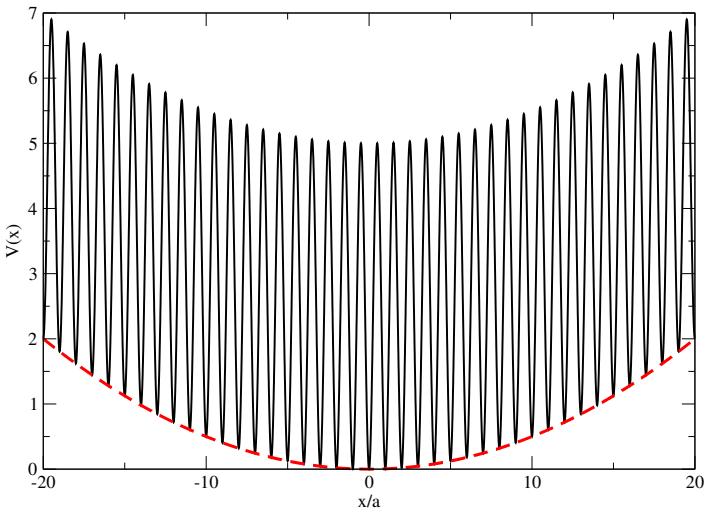

Figure 2.8: Parabolic washboard potential resulting in a logarithmically slow classical annealing. The minima are regularly located at positions $x_i = ia$, and the dashed line shows the parabolic envelope potential.

2.3 One-dimensional curved washboard: a potential with many minima

After discussing at length the annealing problem in a potential with one minimum and with two minima, we wish to move on to a multi-minima problem, however simple. There are simple but interesting one-dimensional potentials which allow us to do that. The first example was proposed and solved by Shinomoto and Kabashima in Ref. [42, 43], and consists in a parabolically shaped washboard potential. This example will display a logarithmically slow classical annealing, showing CA may run into trouble even in simple models with no complexity whatsoever, whereas quantum mechanics can do much better in this case. Consider a wiggly one-dimensional potential with barriers of individual height $\approx B$ separating different local minima, regularly located a distance a apart one from each other, i.e., at positions $x_i = ai$. The ith-local minimum is at energy $\epsilon_i = ka^2i^2/2$, so that the resulting envelope is parabolic. In order to study the dynamics of a particle in this potential, a good starting point is to

write the master equation for the probability $P_i(t)$ that the particle is in the ith-valley at time t:

$$\frac{1}{\gamma}\frac{d}{dt}P_i(t) = P_{i+1}(t)e^{-\beta B_{i+1,i}} + P_{i-1}(t)e^{-\beta B_{i-1,i}} - P_i(t)\left[e^{-\beta B_{i,i+1}} + e^{-\beta B_{i,i-1}}\right] , \tag{2.16}$$

where γ is an attempt frequency, $B_{i,j}$ is the effective barrier from i to j, and $\beta = 1/k_BT$. This is a well justified starting point, in view of the results of the previous sections (Secs. 2.2 and 2.1.1), showing that classical annealing is extremely fast in reducing the width of a probability distribution within each valley to an essentially delta-function like sequence of peaks of strength P_i. The actual form of the effective barriers depends on the way we model the details of the potential, with the only constraint that detailed balance is satisfied, i.e.,

$$B_{i,i+1} - B_{i+1,i} = \epsilon_{i+1} - \epsilon_i = \Delta_i ,$$

in such a way that the stationary solution, for constant T, is simply the Boltzmann probability distribution, $P_i(t \to \infty) \propto exp(-\epsilon_i/k_BT)$. Ref. [42, 43] takes $B_{i+1,i} = B_{i,i-1} = B$, while $B_{i,i+1} = B + \Delta_i$ and $B_{i-1,i} = B + \Delta_{i-1}$, with $\Delta_i = \epsilon_{i+1} - \epsilon_i$. The potential energy of the valleys enters only through the $B_{i,j}$, which control the probability of making transitions between valleys.

In order to study Eq. 2.16, Shinomoto and Kabashima introduced a continuum limit, by defining a macroscopic coordinate x, such that the minima are at $x_i = ia$, and writing the equation governing the probability $P(x,t)$ in the limit $a \to 0$. The derivation involves writing $P_{i\pm1}(t)$ in terms of derivatives of $P(x,t)$, keeping consistently terms up to order a^2 and expanding exponentials with the assumption that $k_BT/(ka^2) >> 1$. The continuum limit equation governing the evolution of $P(x,t)$ turns out to be a Fokker-Planck (FP) equation, Eq. 2.2, with an effective diffusion constant of the form

$$D_{\text{eff}}(T) = \gamma a^2 e^{-B/k_BT} , \tag{2.17}$$

$\eta(T) = k_BT/D_{\text{eff}}(T)$, and an effective drift potential $V(x) = kx^2/2$ given by the macroscopic parabolic envelope potential. In order to study the annealing properties of the system, one can then follow exactly the same steps leading to Eq. 2.4, which applies here too, except that now D_t is substituted by $D_{\text{eff}}(T(t))$, which has an exponential activated behavior, $D_{\text{eff}}(T) \propto e^{-B/k_BT}$. This exponentially activated $D_{\text{eff}}(T)$ changes the annealing behavior in a drastic way. Recall that the CA exponent Ω_{CA} of

Sec. 2.1.1 decreases toward zero as the exponent α_D in the relationship $D(T) \propto T^{\alpha_D}$ increases. Since, close to $T = 0$, $e^{-B/k_BT} << T^{\alpha_D}$ for any arbitrarily large α_D, we could suspect that the behavior of $\epsilon_{res}(\tau)$ will no longer be a power law. In fact, the surprising result of this exercise [42, 43] is that the optimal annealing schedule $T(t)$ is *logarithmic* and that $\epsilon_{res}(t)$ converges to 0 at best as [5]

$$\epsilon_{res}(t) \sim \log(t)^{-1} \ . \tag{2.18}$$

The reason behind this result is that the time derivative of $\epsilon_{res}(t)$ becomes exponentially small as one anneals T toward 0, due to the presence of e^{-B/k_BT} in the diffusion constant, and an exponentially small derivative brings – not surprisingly – a logarithmically slow decrease of the function. To put it more physically, consider solving Eq. 2.4 for a time independent T; the solution is trivially

$$\epsilon_{res}(t) = Ce^{-t/t_{\rm relax}} + \frac{k_BT}{2} \qquad t_{\rm relax} = \frac{k_BT}{2\gamma k a^2} e^{B/k_BT} \ .$$

We observe that the solution converges to the equilibrium (equipartition) value $k_BT/2$ exponentially with a characteristic time $t_{\rm relax}$ which itself increases exponentially fast with decreasing T. As a result, the system will never be able to follow the decreasing T till the end of the annealing, by maintaining roughly the equilibrium value $\epsilon_{pot} = k_BT/2$. Indeed, if we assume for instance $T(t) = T_0(1 - t/\tau)$, the relaxation of the systems will cease to be effective – i.e., the system will fall *out of equilibrium* – at a time t^*, and temperature $T^* = T(t^*)$, at which $t_{\rm relax} \approx \tau$, i.e., when $k_BT^* \approx B/\log\gamma\tau$. The residual energy at this point cannot be smaller then the equipartition value $k_BT^*/2$, hence $\epsilon_{res} \approx B/\log\gamma\tau$ as well. This freezing and falling out of equilibrium for classical systems with barriers seems to provide an ubiquitous source of logarithms in classical annealing [44].

How would one tackle this annealing problem quantum mechanically? As the quantum analog of the master equation Eq. 2.16, we propose studying the Schrödinger

[5] For the sake of precision, we should stress the fact that the logarithm is, strictly speaking, born out of the continuum limit, which is indeed valid only for $T >> \Delta = ka^2/2$. If the lattice constant a is kept finite, there is a minimum non-zero value Δ_V for the splitting between the bottom of any two valleys, and, based on the two-level system results, we could anticipate a final power-law behavior $\tau^{-\Delta_V/B}$. Nevertheless, the logarithmic behavior should apply in a whole temperature window with a final crossover to the power-law behavior.

evolution governed by a tight-binding Hamiltonian with on-site energies ϵ_i and hopping matrix-elements between adjacent sites $h_{i,i+1}$

$$H = \sum_i \epsilon_i c_i^\dagger c_i - \sum_i h_{i,i+1} \left(c_i^\dagger c_{i+1} + c_{i+1}^\dagger c_i \right) \ . \tag{2.19}$$

The justification and possible limitations of this starting point, over the actual original continuum problem will be discussed at the end of the section. Here it is enough to consider that the hopping matrix-elements $h_{i,i+1}$ depend on tunneling through the barrier separating i from $i+1$. The precise form of $h_{i,i+1}$ is likely to be inessential, including its energy (and hence site) dependence, for which we will assume the semi-classical (WKB) form:

$$h_{i,i+1} \sim h = h_0 \left(\frac{V_h}{\Gamma} \right)^{1/4} e^{-\sqrt{V_h/\Gamma}} \ , \tag{2.20}$$

$\Gamma = \hbar^2/(2ma^2)$ being simply related to the quantum confinement energy of a particle of mass m in a valley of size $\sim a$. Here h_0 and V_h are energy parameters related to the details of the potential and of the barrier, which will play little or no role. If the mass of the particle m (and hence Γ and $h_{i,i+1}$) is kept constant the particle will explore the potential due to the kinetic term in the Hamiltonian: the correspondence between the quantum and the classical formulation is that Γ plays the role of T, $h_{i,i+1}$ plays the role of the classical transition probabilities, the ground state wavefunction $|\Psi_i^{(\Gamma,GS)}|^2$ at a given value of Γ (or, equivalently, of the hopping term h) plays the role of the classical equilibrium Boltzmann distribution $P_i^{(T,eq)}$. The question, once again, is how to anneal Γ, by reducing it as a function of time, $\Gamma(t)$, in such a way as to squeeze the wavefunction of the system so that the average potential energy

$$\epsilon_{pot}(t) = \frac{\sum_i \epsilon_i |\Psi_i(t)|^2}{\sum_i |\Psi_i(t)|^2}$$

is minimal.

As it turns out, the continuum limit is once again useful. One goes to the continuum exactly as in the FP case by using $x_i = ai$, writing $\psi(x_i, t) = \Psi_i(t)/\sqrt{a}$ and expanding everything to order a^2. When written in first quantized form, the Hamiltonian for the quantum particle in the macroscopic continuum coordinate x is simply

$$H(t) = -\Gamma_{\text{eff}}(t)\nabla^2 + V(x)$$

where the coefficient of the Laplacian Γ_{eff} is the quantum counterpart of the classical effective diffusion constant D_{eff}

$$\Gamma_{\text{eff}}(t) = a^2 h(t) = a^2 h_0 \left(\frac{V_h}{\Gamma(t)}\right)^{1/4} e^{-\sqrt{V_h/\Gamma(t)}}\,, \tag{2.21}$$

and $V(x) = kx^2/2$, as in the classical case. The continuum limit quantum problem has therefore exactly the form we have considered in Sec. 2.1.1, except that $\Gamma(t)$ in Eq. 2.7 is now substituted by an effective Laplacian coefficient Γ_{eff} which has a highly non-linear, in fact exponential, dependence on the annealing parameter $\Gamma(t)$. We know, however, from Sec. 2.1.1, that a non-linear behavior of the type $(1-t/\tau)^{\alpha_\Gamma}$ for the Laplacian coefficient leads to a power-law decrease of $\epsilon_{res}(\tau)$ with an exponent Ω_{QA} which is, remarkably, an *increasing function of* α_Γ, approaching 1 as $\alpha_\Gamma \to \infty$. Therefore, contrary to the classical case, where an exponentially activated behavior of the diffusion constant D_{eff} is *strongly detrimental* to the annealing (turning a power law into a logarithm), here the exponential WKB-like behavior of Γ_{eff} will do no harm at all. Indeed, we numerically integrated the relevant equation for B_t, Eq. 2.7 with Γ_{eff} in place of Γ, using the exponential WKB expression Eq. 2.21 for Γ_{eff} while annealing $\Gamma(t)$ with a linear schedule, $\Gamma(t) = \Gamma_0(1 - t/\tau)$. The integration was performed, as usual, with a fourth-order Runge-Kutta method, and was carried on up to time $t = \tau$, when the kinetic term in the Hamiltonian ceases to exist. The numerical results (not shown) have a clear power-law behavior for the quantum annealed (QA) final residual energy $\epsilon_{res}(t = \tau) \sim \tau^{-\Omega_{QA}}$, with a power law exponent Ω_{QA} which is compatible with 1. Once again, the exponent appears to be insensitive to the choice of the type of quantum evolution (RT versus IT), although the numerical values of residual energies always respect the inequality $\epsilon^{RT}_{res}(\tau) \geq \epsilon^{IT}_{res}(\tau)$.

Before ending this section, we would like to discuss briefly the reason for treating by tight-binding, Eq. 2.19, what was originally a continuum problem with a well defined potential landscape. As we learned from the double-well case, there is never a clear-cut discrete model (a discrete two-level system, in that case) describing in a complete way the continuum Schrödinger problem, in all stages of the annealing. Obviously, when the mass of the particle is very small, the tight-binding approximation contained in Eq. 2.19 is not particularly good, since more than one state per valley is generally important to describe the wavefunction accurately. As the mass of the particle increases, however, the tight-binding approach gets more and more appropriate, until a further limitation appears: when the mass if very large, it is not legitimate to

neglect excited states within, say, the central valley compared to the lowest states localized in metastable valleys. We can imagine that the ultimate behavior of $\epsilon_{res}(\tau)$, in the quantum case, will be actually dominated by the rather trivial problem of squeezing the wavefunction in the lowest central minimum, with its characteristic power-law exponent ($1/3$, for instance, for a linear schedule). There is, however, an intermediate region, between the very short τ scale, where the full details of the potential are important, and the very long τ scale, where the trivial squeezing mentioned above sets in, and where the tight-binding approximation reasonably predicts a power-law exponent (of order 1) for $\epsilon_{res}(\tau)$.

We believe that one of the important points that makes QA so different from CA in the present case is that the spectrum of the instantaneous eigenvalues of the quantum problem does not show any dangerous Landau-Zener avoided-crossing, and, correspondingly, the ground state wavefunction is always more peaked in the central valley than elsewhere. As in the two-level case, a disorder in the width of the different valleys would change this result.

2.4 Role of disorder

Despite their disarming simplicity, the three case studies above turn out to be extremely informative in qualifying the profound difference of QA from CA, and their surprising consequences. We expect that these results will be very important in understanding more realistic QA problems. Of course, the cases studied, although instructive, do not possess the real ingredient which makes annealing difficult, both in CA and QA, i.e., some form of disorder in the distribution of the minima. We believe, for instance, that even an irregular landscape with many minima, as the double-cosine potential $V(x) = V_1 \cos(2\pi x) + V_2 \cos(2\pi r x)$ (with r an irrational number) shown in Fig. 2.9, would already change drastically the behavior of QA from a power-law to a logarithm. On quite general grounds, Anderson's localization [47] would predict that wavefunctions are localized for a genuinely disordered potential for large enough mass (i.e., small enough kinetic energy bandwidth) in any $D > 2$ (this localization occurs for all value of the mass in $D = 1, 2$). Therefore, quantum annealing should always, via a cascade of Landau-Zener events, end up into some localized state which has, *a priori*, nothing to do with our search of the actual potential minimum.

A very simply illustration of the crucial role of disorder is given by the $D = 1$

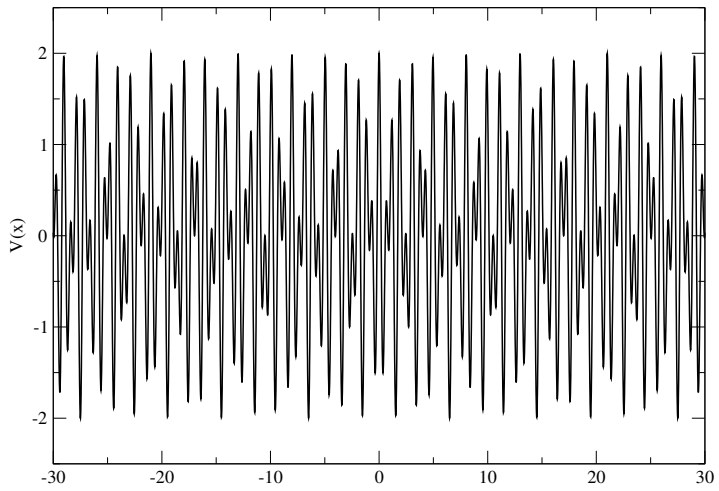

Figure 2.9: Double cosine potential $V(x) = cos(2\pi x) + cos((1+\sqrt{5})\pi x)$, showing an irregular landscape with many minima.

disordered Ising ferromagnet:

$$H = -\sum_i J_i \sigma_i^z \sigma_j^z - \Gamma \sum_i \sigma_i^x$$

where $J_i \geq 0$ are positive random variables in the interval $[0, 1]$, and Γ is the transverse field inducing quantum fluctuations. Obviously, the ground state is the ferromagnetic state with all spins aligned up (or down). However, arbitrarily weak values of the J_i can pin domain walls between up and down ferromagnetic regions, with a very small energy cost $2J_i$. For a finite system with periodic boundary conditions domain walls appear in pairs, and separate sections of the system with alternating $\uparrow$ and $\downarrow$ ferromagnetic ground states. Given two domain walls pinned at weak-J_i points a distance $L \gg 1$ apart, healing the system via single spin flip moves requires flipping L spins, which can be a formidable barrier to tunnel through. The system will have a *very slow annealing* (quantum, as well as classical) while showing, at the same time, *no complexity* whatsoever: simple disorder is enough.

Chapter 3

Annealing by Classical Monte Carlo

In the previous chapter we studied the annealing dynamics of a simple double-well system, by means of an exact integration of the dynamical equations of motion: Fokker-Planck, for the classical case, or Schrödinger, for the quantum case. Unfortunately, such a direct deterministic approach is applicable, for practical purposes, only to small dimensional optimization problems. To appreciate the difficulty, it is enough to think that the number of possible configurations, and hence the Hilbert space, of a small 32×32 square lattice Ising model is $2^{1024} \sim 10^{308}$, an astronomically large number which forbids a direct deterministic dynamical approach. Since the hard instances of typical optimization problems involve an even larger number of variables, there is an obvious need for an alternative strategy.

As first pointed out by Kirkpatrick *et al.* in their first application of simulated annealing [1], a classical Monte Carlo scheme can be used to overcome this difficulty. More recently, Quantum Monte Carlo algorithms have been implemented as a tool for performing quantum annealing simulations [2, 8, 9, 10, 12, 13].

We recall that by "Monte Carlo" we simply mean a way of implementing a stochastic (rather than exact) dynamics (a *Markov chain*), by making use of a (pseudo) random number generator [48]. However, it is worth observing that deterministic dynamics, FP or Schrödinger, and Monte Carlo, classical or quantum, are not equivalent, i.e. they do not provide the same dynamics: there is no general relationship between the (physical) time appearing in the Fokker-Planck or Schrödinger equation, and the corresponding classical or quantum Monte Carlo time-step. As a consequence, many different Monte Carlo dynamics are possible, even if they converge to the same equi-

librium probability distribution [48].

In this chapter, we shall first give a general introduction to Markov processes and classical Monte Carlo (MC), and then discuss the results of a set of possible implementations of MC dynamics for classical annealing of the double well system. Much of the general framework we will illustrate applies to a general problem in a (discrete or continuous) multidimensional configuration space, whose generic point we denote by x. The detailed Monte Carlo results, however, will refer to the double-well potential of the previous chapter. In the final part of the chapter, we will provide a theoretical explanation of the MC results from a spectral analysis, both numerical and analytical, of the relevant Markov transition matrix.

3.1 Markov processes

Suppose we have a system which is described by a certain configuration space x, possibly multidimensional. The configuration space can be a continuum (like in our double-well potential), or it can be discrete (for instance, for an Ising problem, x is just the configuration $\{S\} = (S_1, \cdots, S_N)$ of all the N spins, i.e., a discrete-value variable in a 2^N-dimensional space). For simplicity of notation, we will assume the x-space to be discrete: a continuum case is recovered, as usual, by changing all the sums into integrations.

A *Markov process* (or Markov chain) in x-space is a *non-deterministic discrete-time dynamics*, for which the probability of having a configuration x_n at a certain time $t_n = n\Delta t$, denoted by $P_n(x_n)$, evolves, as a function of a discrete iteration time t_n, according to a *Master equation* of the form:

$$P_{n+1}(x_{n+1}) = \sum_{x_n} W(x_{n+1}, x_n)\, P_n(x_n) \ , \tag{3.1}$$

where $W(x_{n+1}, x_n)$ is the *conditional probability* (or transition operator) for moving from x_n to x_{n+1} in the given time-step Δt. The conditional probability, which does not depend on n due to time homogeneity, is a *non-negative* and *column-normalized* (generally non-symmetric) matrix, i.e.,

$$\begin{aligned} & W(x', x) \geq 0 \\ & \sum_{x'} W(x', x) = 1 \ . \end{aligned} \tag{3.2}$$

A distinctive feature of a Markov process is that P_{n+1} depends only on P_n, and not on previous times.

The Master equation allows us, in principle, to calculate, iteratively, the time-evolution of the probability P_n, once an initial condition is given, for instance at iteration $n = 0$, through a $P_0(x_0)$. We stress that, although the actual value of the random variable x_n at iteration n is not known deterministically (the Markov dynamics is inherently stochastic), its probability distribution $P_n(x_n)$ is instead known *exactly*, in principle, if the Master equation is solved.

Despite the fact that a general solution of the Master equation is in general not known, many useful properties can be derived, and a quite general theory is available for such a mathematical object [41]. First of all, a Markov process is convergent, $P_n(x) \to P_{eq}(x)$ for $n \to \infty$, under quite general circumstances, to a unique equilibrium distribution (see Ref. [49][page 53ff] for a complete proof). A sufficient condition for this to occur, is that the transition operator W is *ergodic* (i.e., starting from a configuration x there is a non-vanishing probability of reaching any other configuration x' in a finite number of Markov time steps) and that is satisfies a *detailed balance* condition:

$$W(x', x)\, \bar{P}(x) = W(x, x')\, \bar{P}(x') \ , \tag{3.3}$$

with some $\bar{P}(x) \geq 0$. [1] $\bar{P}(x)$ appearing in Eq. 3.3 is exactly the unique stationary solution to which the Master equation converges for $n \to \infty$, $P_n(x) \to P_{eq}(x) = \bar{P}(x)$.

It is possible to take the *time-continuum limit* of the previous discrete-time dynamics: this will be useful later on, in discussing the spectral properties of the Master equation. Denote, for simplicity, $P(x, t_n) = P_n(x)$, and rewrite the Master equation 3.1 as a finite difference:

$$\frac{P(x', t_n + \Delta t) - P(x', t_n)}{\Delta t} = \sum_{x \neq x'} \frac{W(x', x)}{\Delta t} P(x, t_n) - \sum_{x \neq x'} \frac{W(x, x')}{\Delta t} P(x', t_n) \ . \tag{3.4}$$

In deriving Eq. 3.4 we have made use of the column normalization of W ($\sum_x W(x, x') = 1$), and we have got rid of the $x = x'$ contributions in the two sums, because they can-

[1] This relationship indicates that the number of processes undergoing a transition $x \to x'$ has to be exactly compensated, to maintain a stable stationary condition, by exactly the same amount of reverse processes $x' \to x$.

cel. In the limit $\Delta t \to 0$, is is quite natural to postulate [41] that

$$W(x',x) = e^{\Delta t \tilde{W}} = \delta_{x',x} + \Delta t \, \tilde{W}(x',x) + \cdots \tag{3.5}$$

where $\tilde{W}(x',x)$ is a *probability per unit time* of making a transition from x to x'. Obviously, the conditions on W in Eq. 3.2 imply that, to leading order in Δt:

$$\begin{aligned} &\tilde{W}(x',x) \geq 0 \qquad \text{for} \quad x' \neq x \\ &\sum_{x'} \tilde{W}(x',x) = 0 \; . \end{aligned} \tag{3.6}$$

In terms of $\tilde{W}$, it is immediate to take the limit $\Delta t \to 0$ of Eq. 3.4, obtaining:

$$\frac{\partial P(x',t)}{\partial t} = \sum_{x} \tilde{W}(x',x) \, P(x,t) \; . \tag{3.7}$$

Notice, finally, that neither W nor $\tilde{W}$ are, in general, symmetric matrices. However, if detailed balance holds, they can be readily symmetrized. This symmetrization is the precise analogue of the mapping of the Fokker-Planck equation onto an imaginary-time Schrödinger problem (see App. A.2). Indeed, by dividing both terms of the detailed balance expression in Eq. 3.3 by $(\Delta t)\sqrt{\bar{P}(x)\bar{P}(x')}$, were $\bar{P} = P_{eq}$, we immediately see that the matrix

$$\bar{H}_{x',x} = -\frac{1}{\sqrt{\bar{P}(x')}} \tilde{W}(x',x) \sqrt{\bar{P}(x)} = \bar{H}_{x,x'} \; , \tag{3.8}$$

is symmetric. As a by-product – using the above equation and Eq. 3.5 – we also obtain that:

$$W(x',x) = \sqrt{\bar{P}(x')} \, e^{-\Delta t \, \bar{H}}(x',x) \, \frac{1}{\sqrt{\bar{P}(x)}} \; . \tag{3.9}$$

If we introduce the (wave)function $\phi(x,t)$ such that $P(x,t) = \sqrt{\bar{P}(x)}\psi(x,t)$, we can rewrite the Master equation Eq. 3.7 as an imaginary-time Schrödinger problem which has $\bar{H}$ as Hamiltonian:

$$-\frac{\partial \psi(x',t)}{\partial t} = \sum_{x} \bar{H}_{x',x} \, \psi(x,t) \; . \tag{3.10}$$

The minus sign introduced in the definition of $\bar{H}$ is merely a matter of convenience; In this way: i) the imaginary-time Schrödinger problem has the conventional sign,

and ii) the *off-diagonal* matrix elements of $\bar{H}$ are *non-positive* (see Eq. 3.6), so that the Perron-Frobenius theorem [49][page 54ff] – which guarantees the uniqueness and positiveness of the ground state – applies. To finish this rather general section, it is a very simple matter to verify that $\phi_0(x) = \sqrt{\bar{P}(x)}$ is a positive eigenfunction of $\bar{H}$ with eigenvalue $\lambda_0 = 0$. By the Perron-Frobenius theorem, ϕ_0 must then be the unique ground state of $\bar{H}$, and every other eigenvalue of $\bar{H}$ must be positive $\lambda_{i>0} > 0$.

Summarizing, if a Markov chain satisfies detailed balance and is ergodic, then an equilibrium distribution $\bar{P}(x)$ will always be reached, for large enough n, independently of the initial condition $P_0(x)$ at $n = 0$. [2] The convergence for large n (or t, in the time-continuum formalism) is guaranteed to be exponentially fast, and is governed by $e^{-\lambda_1 t}$ where λ_1 is the smallest non-zero eigenvalue of the symmetric matrix $\bar{H}$ associated to the transition operator W.

3.2 Metropolis algorithm

We now turn to the actual *construction* of a Monte Carlo (MC) algorithm. For that purpose, we want to construct a Markov chain (i.e., a $W(x', x)$ satisfying Eq. 3.2) such that, for large n, the configurations x_n generated are distributed according to a given probability distribution, $\bar{P}(x) = P_{eq}(x)$: for instance, in the present context, the Boltzmann probability $P_{\text{Boltzmann}}(x)$. We are free to choose $W(x', x)$ as we wish, and therefore insist that W verifies the detailed balance condition, i.e.,

$$W(x', x)\, P_{eq}(x) = W(x, x')\, P_{eq}(x') \; . \tag{3.11}$$

How do we do that, in practice? There is a quite large freedom in the choice of W, and the construction comes therefore in many flavors (Metropolis MC, Heat Bath MC, etc.). Metropolis and collaborators [50], for instance, introduced the following very simple scheme. They started considering a trial (attempt) transition probability $T(x', x)$, defining the probability of going from x to x', which can be chosen with great freedom, without any requirement of detailed balance. Basically, a trial distribution T must be: i) simple to implement on a computer, ii) ergodic (any configuration must be reachable in a finite number of steps), and iii) non-negative ($T(x', x) \geq 0$) and normalized to one ($\sum_{x'} T(x', x) = 1$). Many simple examples are possible: for

[2] $\bar{P}(x)$ is the distribution appearing in the detailed balance equation.

a continuum problem, for instance, one can take a flat distribution on a compact domain (Box), a multidimensional Gaussian, or a Lorentzian (Cauchy) distribution (see Sec. 3.3 for details). In order to define a $W(x', x)$ satisfying the detailed balance condition in Eq. 3.11, the new configuration x' generated by the chosen trial transition probability $T(x', x)$ has to be *accepted* only with a probability:

$$A(x', x) = \min \left\{ 1, \frac{T(x, x') P_{eq}(x')}{T(x', x) P_{eq}(x)} \right\} , \tag{3.12}$$

so that the resulting conditional probability $W(x', x)$ is given by:

$$W(x', x) = A(x', x)\, T(x', x) \quad \text{for } x' \neq x . \tag{3.13}$$

(The value of $W(x, x)$ is fixed by the normalization condition, $\sum_{x'} W(x', x) = 1$.) The proof that detailed balance is satisfied by the $W(x', x)$ in Eq. 3.13 is quite elementary, and can be found in Ref. [51].

Summarizing, if x_n is a configuration at time t_n, the Markov chain iteration, from t_n to t_{n+1}, is defined in two steps:

1. A *move* is proposed by generating a configuration x' according to the trial transition probability $T(x', x_n)$;

2. The move is *accepted*, and the new configuration x_{n+1} is taken to be equal to x', with probability $A(x', x_n)$ (practically, a random number ξ_n uniformly distributed in the interval $(0, 1]$ is extracted, and the move is accepted if $\xi_n \leq A(x', x_n)$), otherwise the move is *rejected* and one keeps $x_{n+1} = x_n$.

Such a process (trial move followed by an acceptance/rejection step) is the basic *Monte Carlo step*, and must be iterated repeatedly.

The length of this MC time series must be carefully chosen. First of all, the equilibrium distribution is reached only after a suitable number of MC steps $n > N_{\text{equilibration}}$, depending on the equilibration time of the process, which could be very long. [3] Therefore, average quantities have to be calculated only after having skipped of the order of $N_{\text{equilibration}}$ steps. Secondly, since only a finite number of configurations can be actually generated, one must take into account the statistical error due to

[3] Recall that the leading correction to the $n \to \infty$ limit is governed by $e^{-\lambda_1 t}$, where λ_1 is the smallest non-zero eigenvalue of the symmetric matrix $\bar{H}$ associated to W (see Sec. 3.1).

the truncation of the MC series. Consider the mean value of a given (local) observable $O(x)$, which can be estimated by:

$$\bar{O} = \frac{1}{N} \sum_{i=1}^{N} O(x_i) \; .$$

Quite generally, two configurations x_i and x_j will be correlated unless they are separated by a suitably long time-interval, of the order of the so-called *correlation time*. [4] The computation of the truncation error involves explicitly the correlation length. Indeed, the equation $\mathrm{var}(\bar{O}) = \frac{\mathrm{var}(O)}{N}$ is valid only if the sampled configurations x_i are completely uncorrelated. [5] In more general cases, one must make use of the equation:

$$\mathrm{var}(\bar{O}) = \frac{var(O)}{N} \; (1 + 2\,\tau_{cor}) \, , \tag{3.14}$$

where τ_{cor} is the correlation time [53, page 69], which also appears in the so-called *autocorrelation function*:

$$C(t) = \frac{1}{N-t} \sum_{i=1}^{N-t} x(i+t)\, x(i) - \bar{x}^2 \sim e^{-\frac{t}{\tau_{corr}}} .$$

For a more complete discussion, we refer the reader to Ref. [52].

3.3 Different trial moves for the double-well potential

As stressed in the previous section, the setting up of a MC algorithm to sample a given equilibrium distribution P_{eq} gives a large amount of freedom in the construction of the transition operator W. Even within the Metropolis algorithm, we still have a large freedom in choosing the trial transition probability $T(x', x)$.

Let us restrict now our attention to the double-well potential case. The configuration space x is here the real line.

[4] We stress again that time here has to be understood as number of Monte Carlo steps, i.e., a trial move followed by an acceptance process.

[5] The variance of the observable O can be estimated through the formula: $\mathrm{var}(O) = \frac{1}{N} \sum_{i=1}^{N} O^2(x_i) - \bar{O}^2$, while the variance of the *mean* of the same observable is given by: $\mathrm{var}(\bar{O}) = \langle \bar{O}^2 \rangle - \langle \bar{O} \rangle^2$, where the angle brackets denote an average over different realizations of the same Markov chain (see [52]).

A first obvious choice for $T(x',x)$ is given by the *Box move* (historically, the first one used [50]):

$$T_{Box}(x',x) = \begin{cases} \frac{1}{2\sigma} & \text{if} \quad |x'-x| < \sigma \\ 0 & \text{otherwise} \end{cases}$$

where the length parameter σ controls the move range. According to the Metropolis prescription, see Eq. 3.13, the transition operator is written as:

$$W_{Box}(x',x) = \begin{cases} \frac{1}{2\sigma} \min\left(1, \frac{P_{eq}(x')}{P_{eq}(x)}\right) & \text{if} \quad 0 < |x'-x| < \sigma \\ 0 & \text{if} \quad |x'-x| \geq \sigma \end{cases} \tag{3.15}$$

An alternative, quite often used, choice for $T(x',x)$, is given by the *Gaussian move*:

$$T_{Gau}(x',x) = \frac{1}{\sqrt{2\,\pi}\,\sigma} \exp\left(-\frac{(x'-x)^2}{2\,\sigma^2}\right) ,$$

which results in:

$$W_{Gau}(x',x) = \frac{1}{\sqrt{2\,\pi}\,\sigma} \exp\left(-\frac{(x'-x)^2}{2\,\sigma^2}\right) \min\left(1, \frac{P_{eq}(x')}{P_{eq}(x)}\right) \quad \text{if} \quad x' \neq x \; . \tag{3.16}$$

Finally, one can also introduce a *Lorentzian move*:

$$T_{Lor}(x',x) = \frac{1}{\pi} \frac{\sigma}{(x'-x)^2+\sigma^2} ,$$

with a corresponding W given by:

$$W_{Lor}(x',x) = \frac{1}{\pi} \frac{\sigma}{(x'-x)^2+\sigma^2} \min\left(1, \frac{P_{eq}(x')}{P_{eq}(x)}\right) \quad \text{if} \quad x' \neq x \; . \tag{3.17}$$

In all cases, σ controls the move range. We also emphasize that the dependence of $W(x',x)$ on $\Delta\,t$ is now hidden in the σ, which is the only free parameter. [6] We note that all the trial moves T (Box, Gaussian, and Lorentzian) reduce to a Dirac's delta when $\sigma \to 0$. Nevertheless, the three choices are very different: the Box distribution has a compact support, while the two others have (long) tails extending over the whole space. Moreover, the Gaussian and the Lorentzian differ in that the *second moment* of the former is finite, while that of the latter diverges. We will see below that while, by construction, all these W sample the same equilibrium distribution $P_{eq}(x)$, the *Markov dynamics* associated to them are *different*, leading to different annealing behaviors of the corresponding MC algorithms.

[6] This is a natural assumption for every diffusive process: the longer is the process time the wider will be the spreading of the initial distribution. In particular, it is possible to derive the analytical time-dependence of the move range $\sigma(\Delta\,t)$ by means of the exponential map, Eq. 3.5.

3.4 Derivation of the Fokker-Planck equation

Before illustrating the actual MC results, let us pause for a second to explore the relationship between the continuous-time Master equation, Eq. 3.7, with a certain choice of transition operator $\tilde{W}$ (see Sec. 3.3), and the Fokker-Planck equation studied in Chap. 2.

For σ "small enough" (with respect to the typical length scales of the potential) it must be possible to expand $\tilde{W}(x', x)$ in Eq. 3.7 up to the second order in $|x' - x|$ (see Ref. [41][page 214ff]). If we do so, we obtain that the Master Eq. 3.7 reduces to the following Fokker-Planck equation:

$$\frac{\partial P(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[a_1(x)\, P(x,t)\right] + \frac{1}{2}\left[a_2(x)\, P(x,t)\right], \tag{3.18}$$

where

$$a_k(x) = \int \mathrm{d}y\, (y-x)^k\, \tilde{W}(y,x)$$

are the so-called *jump moments* of order k. These are the first two terms of an expansion known as Kramers-Moyal series [54][page 63ff]. According to a theorem due to Pawula [54][page 70ff], such a second-order approximation holds under quite wide conditions.

On the basis of this small σ limit, we expect that the Metropolis dynamics should reduce, apart from a generally unknown time-scale, to the previously studied Fokker-Planck dynamics if the conditions are such that the small σ limit dominates the dynamics. These conditions will be met for the Box trial move, as we shall see.

3.5 Classical Monte Carlo simulations on the double-well potential

In this section we shall report the results of a classical annealing of the double-well potential, implemented by means of Metropolis MC employing the three types of trial moves illustrated in Sec. 3.3. Before illustrating the results, we need to specify a few more technical details.

As in Chap. 2, we would like to start annealing from an equilibrium distribution at a given (relatively high) temperature T_0. This is easily done, within MC, by performing, before the actual annealing, a (short) equilibrium MC run at a given fixed

temperature T_0 (equilibrium is reached after a short equilibration time at high T). After this initial equilibration part, the actual annealing starts. In order to compare with the Fokker-Planck results of Chap. 2, we set $T_0 = V_0$ and we employ a linear annealing schedule: $T(t) = T_0\ (1 - t/\tau)$ (see Sec. 2.1). The internal time t and the annealing time τ are now both measured in units of MC steps. In order to calculate the statistical error of a given quantity, however, we cannot use the technique, adopted in equilibrium MC, of making very long simulations to accumulate uncorrelated data. In our out-of-equilibrium case, the statistical error of an observable must be calculated by performing several independent equilibration-plus-annealing runs, for each given value of τ.

One last concern regards the choice of the length σ parameterizing the range of the trial move (see Sec. 3.3), and its behavior as a function of t during annealing. σ is strictly correlated with the *root mean square (r.m.s.) displacement* of x

$$\tilde{D} = \langle (x(t + \Delta t) - x(t))^2 \rangle \ .$$

(We employed $\tilde{D}$ instead of D in order to distinguish the r.m.s. displacement from the diffusion constant.) At any given fixed T, we carried out several equilibrium MC runs at different values of σ, with the goal of *maximizing* the r.m.s. displacement $\tilde{D}$. [7] This procedure fixes, for any given T, the σ value which maximizes $\tilde{D}$. In Fig. 3.1 we plot the resulting $\tilde{D}$ (top) and σ (bottom) versus the temperature T, for equilibrium MC simulations with the Box move. The values plotted are also reported in Table 3.1. It is worth observing (see Table 3.1) that the value of σ which maximizes $\tilde{D}$ leads also to an average acceptance close to 50%, which is a well-known rule-of-thumb for optimal sampling in equilibrium MC simulations. Fig. 3.1 shows also power-law interpolations for both the quantities shown. Admittedly, the results are a bit scattered, but the approximation $\tilde{D} \propto T$ and $\sigma \propto T^{1/2}$ seems to catch the main data trend. There is also a simple physical reason behind such a behavior. Indeed, $\tilde{D}$ is related to the diffusion constant D appearing in the Fokker-Planck equation $D = \tilde{D}/(2\Delta t)$ (in the short-time approximation, $\Delta\, t \to 0$), and one expects that the *Einstein relation* $D \propto T$ (originally stated for the Brownian motion, and nowadays understood in a more general context of the Langevin equation theory) holds [41][page 217ff]. Moreover, being σ the natural length-scale of the displacement, we expect $\tilde{D} \propto \sigma^2 \propto T$.

[7] By equilibrium simulation we mean that the temperature was held fixed during the simulation.

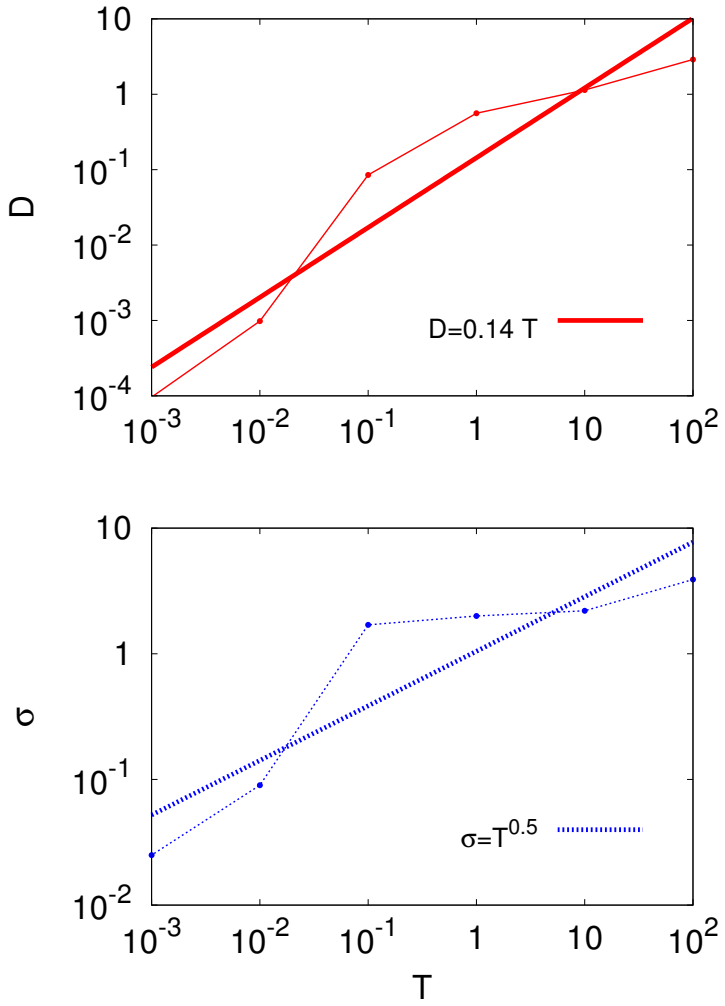

Figure 3.1: Optimal r.m.s. displacement $\tilde{D}$ and parameter σ versus T for equilibrium MC simulations with the Box trial move. The straights lines are power-law interpolations.

Summarizing, it seems natural to impose a square-root schedule for the σ, i.e. $\sigma(t) = \sigma_0 \sqrt{T(t)/T_0}$, and the only parameter we need to fix is the initial condition σ_0. From Fig. 3.1, we see that $\sigma_0 = 2.0$ for the Box trial move. A similar procedure has been carried out for both the Gaussian and the Lorentzian moves, obtaining, respectively, $\sigma_0 = 2.0$, and $\sigma_0 = 2.9$.

We finally turn to illustrating the actual annealing results. Fig. (3.2) shows the average residual energy $\epsilon_{res}(\tau)$ at the end of the annealing (see Sec. 2.1) for the three possible choices of trial move (Box, Gaussian, Lorentzian). The average $\epsilon_{res}(\tau)$ and its error-bar are calculated using 10^5 independent runs for each annealing time τ. The

T	σ	$\tilde{D}$	Acceptance
100	3.9	2.89	54%
10	2.2	1.14	57%
1	2.0	0.56	43%
0.1	1.7	0.085	11%
0.01	0.09	$9.9\,10^{-4}$	43%
0.001	0.025	$9.6\,10^{-5}$	47%

Table 3.1: Optimal r.m.s. displacement $\tilde{D}$, parameter σ, and average acceptance ration (see text) for a set of equilibrium MC simulations with the Box trial move, at several temperatures T.

upper panel of Fig. (3.2) refers to the symmetric double-well potential (SYM), while the lower one refers to the asymmetric one (ASYM, see Sec. 2.2 for their definition). As a reference, we also plot the results obtained by means of the numerical integration of the Fokker-Planck equation (see Sec. 2.2). [8] First of all, we see that the differences between the two potential cases – SYM and ASYM – are negligible. We conclude that, as expected, the actual shape of the potential well does not affect the annealing dynamics, except for the barrier height. Moreover, the Box trial move leads to results in close agreement with those obtained by means of the exact integration of the FP equation. Although the almost perfect superposition of the data is accidental, because the overall MC time-scale is undetermined, the asymptotic behavior at large τ is not accidental. [9] Nevertheless, as discussed in Sec. 3.4, in the limit $\sigma \to 0$ the Markov dynamics associated to the MC should reduce to a form of Fokker-Planck dynamics, at least in the case of a Box trial move, where a small σ strictly cuts off large displacements $|x' - x|$. (See App. B for a more detailed study of this limit.)

Quite evidently, both the Gaussian and the Lorentzian trial moves allow to reach residual energy $\epsilon_{res}(\tau)$ much below that of the Box move (and of the FP dynamics). The Lorentzian case is particularly impressive. In Table 3.2, we report the annealing

[8] We stress that the MC annealing time τ and that of the FP dynamics are not comparable, and it would be dangerous to compare anything but the power law behavior in the asymptotic region. In particular, $\tau = 1$ in the MC case means that a single MC step has been performed, and, for all practical purposes, the configurations are still distributed according to the initial equilibrium distribution.

[9] In general, there is no reason to expect even a linear scaling between exact integration τ and MC τ.

exponent Ω_{CA}, derived by fitting the asymptotic part of the data $\epsilon_{res}(\tau) \propto \tau^{-\Omega_{CA}}$.

	Ω_{CA}
Fokker-Plank	0.14
Box	0.13
Gaussian	0.31
Lorentzian	1.0

Table 3.2: Annealing exponents (see text) obtained by means of exact integration (Fokker-Planck) and MC algorithm using several proposal move (Box, Gaussian, Lorentzian). These data refer to the symmetric (SYM) double-well potential.

A brief discussion is here in order. The fact that the Gaussian trial move is more effective than the Box one is understandable if one considers that the former always allows, even at small T, for long "jumps" over the barrier. What is less obvious, from this point of view, is why the Lorentzian case, which also allows similar long "jumps", is so different, and much more effective, than the Gaussian. This could be an artifact of the simple potential landscape we considered: it could be that for a much rougher potential landscape all the MC trial moves could provide comparable results. We point out, however, that Ref. [55] reports a genuine improvement of the sampling by making use of the Lorentzian distribution, instead of the Gaussian, for non-trivial problems. Moreover, in Ref. [56] a good performance of Lorentzian trial move is reported for the J.J. Thompson problem (finding the equilibrium distribution of N charged particles on a sphere), and in optimizing the structure of a cluster made of N Ni atoms. The authors Ref. [56] suggest also that more exotic distributions, like the non-extensive generalization of the Boltzmann distribution [57], might help the relaxation dynamics of a classical annealing MC. These speculations apart, the result that a Lorentzian move is often more effective than a Gaussian is apparently true even for less-trivial optimization problems. In the remaining part of this chapter we shall put this insight on a more sound mathematical ground, at least for the double-well case we are considering.

3.6 Spectral analysis of a Markov process

The results presented in the previous section show that the choice of the trial move $T(x', x)$ in the Metropolis MC scheme strongly affects the classical annealing behavior for a double-well potential. We speculated that different moves likely lead to very different relaxation times. In the remaining part of this chapter we shall prove this statement by means of a quantitative analysis.

The idea is very simple. For a fixed temperature T, the Markov process admits a spectral analysis: In continuous time, the result is very simple, since we can borrow it from the imaginary-time Schrödinger form of Eq. 3.10. [10] It is straightforward to verify that the expression for the probability distribution at time t is:

$$P(x,t) = P_{eq}(x,t) + \sum_{i>0} a_i \, e^{-\lambda_i t} \, P_i(x) \; , \tag{3.19}$$

where $\{\lambda_i\}_{i>0}$ are the positive eigenvalues of $-\tilde{W}$ (and of $\bar{H}$, see Sec. 3.1), while $P_i(x) = \phi_0(x)\phi_i(x) = \sqrt{P_{eq}(x)}\phi_i(x)$, $\{\phi_i\}$ being the eigenvectors of $\bar{H}$. The coefficients a_i depend on the initial condition. (Since $P_i(x)$ is the i-th right-eigenvector of $-\tilde{W}(x', x)$, it follows that it will be also the right-eigenvector of $W(x', x)$ (see Eq. 3.5), whose corresponding eigenvalue is $\bar{\lambda}_i = e^{-\Delta t \lambda_i}$.) Eq. 3.19 shows that the approach to the equilibrium distribution is governed by a *relaxation time* $t_{rel} = \lambda_1^{-1}$, λ_1 being the smallest non-zero eigenvalue of $-\tilde{W}$ (recall that $\lambda_0 = 0$). [11] Therefore, the largest is the *spectral gap* $\Delta = \lambda_1 - \lambda_0 = \lambda_1$, the fastest the system reaches equilibrium.

How can we use this result in an annealing context? It is natural to assume that, for a given schedule $T(t)$, the best we can do is to *maximize* the spectral gap for each given temperature $T(t)$. This is indeed what we will explore in the following. Generally speaking, finding explicitly the spectral gap Δ for a given $\tilde{W}$ is an impossible task; in the present one-dimensional double-well context, however, it turns out that the spectrum of W is quite simple to obtain by discretizing the real axis. [12]

[10] A similar analysis can be carried out for the original discrete-time Markov chain, see Ref. [41].

[11] Notice that, as a further consequence of the exponential mapping, $\lambda_1 = 1 - \bar{\lambda}_1 + 0\,(\Delta t)$.

[12] One can feel uneasy since in the definition of $W(x', x)$ an undefined constant Δt remains, which seems to spoil the generality of the results. However in the common Metropolis MC practice this constant time-scale is disregarded, and only the proposal move range σ is taken into account (see Sec. 3.3). In particular, σ is the parameter one can tune in order to obtain better convergence of the

3.7 Numerical diagonalization of W

In this section we shall apply the spectral analysis to the proposal moves we defined in Sec. 3.3. We recall that they are: the Box move, see Eq. 3.15, the Gaussian move, see Eq. 3.16, and the Lorentzian one, see Eq. 3.17. The main quantity under inspection is the *spectral gap* $\Delta(T,\sigma) = 1 - \bar{\lambda}_1(T,\sigma)$ which is a function of both the temperature T and the proposal move range σ. We obtain this gap (as well as the higher eigenvalues) by diagonalizing a discretized version of the transition operator $W(x',x)$. [13]

In Fig. 3.3 we plot the gap Δ versus σ at two fixed temperatures, $T = 0.1$ (Top) and $T = 0.001$ (Bottom). As a general remark, notice how the behavior of $\Delta(\sigma)$ for different trial moves is remarkably different. Interestingly, for every proposal move and every T, there is a value of σ which maximizes the gap Δ. Recalling that the relaxation time is proportional to Δ^{-1} we anticipate that, at every temperature T, a $\sigma(T)$ exists which provide the most rapid relaxation dynamics. Clearly this $\sigma(T)$ is *a priori* a better choice than taking just $\sigma(T) = \sigma_0 \, (T/T_0)^{1/2}$, as previously implemented in Sec. 3.5. Notice, however, that at low T the dependence of Δ on σ is no longer smooth. In particular, for $T = 0.001$, the Box trial move shows a sharp transition from vanishingly small gap values to finite ones, while both Gaussian and Lorentzian moves show a cusp maximum. At low temperatures, the maximum gap $\Delta_{max}^{(Lor)}$ of the Lorentzian move is bigger than $\Delta_{max}^{(Gau)}$, which is in turn bigger than $\Delta_{max}^{(Box)}$; correspondingly, for small σ, we have $\Delta^{(Lor)}(\sigma) > \Delta^{(Gau)}(\sigma) > \Delta^{(Box)}(\sigma)$. This is, essentially, the reason for the faster relaxation dynamics of the Lorentzian move during annealing (i.e., varying T, with the same $\sigma(T)$ for all move choices). Notice also that the maximum value of Δ, Δ_{max} decreases with temperature in all cases, as one can realize by comparing the two panels in Fig. 3.3. Moreover the value of σ that maximizes the Lorentzian and Gaussian gap decreases with temperature, while it seems to converge to a finite value (around $\sigma = 1.75$) for the Box case.

The Box case is particularly intriguing, and deserves a closer inspection. In Fig. 3.4 we plot Δ versus the inverse temperature $1/T$ for the case of the Box move,

MC simulation (see Secs. 3.5 and 3.8). Moreover, the $W(x',x')$ defined in Sec. 3.3 are well-defined operators, while their corresponding $\tilde{W}(x',x)$ are singular, as one can verify by means of inverse Fourier transformation (the Gaussian and Lorentzian cases are particularly simple).

[13] We made use of a uniform grid of up to 10^3 points in the interval $[-3,+3]$ of the x-space, just as in Sec. 2.2. This discretization appears to be suitable for the low-lying part of the spectrum we want to study.

and for several values of σ. For a whole range of σ, $\sigma < \sigma_{cr} \approx 1.75$, we see a low-temperature behavior typical of an activated Arrhenius process $\Delta(T) \propto e^{-B_{eff}(\sigma)/T}$, where $B_{eff}(\sigma)$ is an *effective barrier*, clearly σ-dependent, which the system experiences. Above a certain σ_{cr}, the behavior of Δ changes drastically, from Arrhenius to what appears, at first sight, just a constant. A closer inspection on an extended temperature range, see Fig. 3.5, shows that at very small temperatures an (avoided) crossing between $\Delta = 1 - \bar{\lambda}_1$ and $1 - \bar{\lambda}_2$ occurs, and Δ starts to decrease again toward zero as $\Delta \propto T^{1/2}$ (see App. B.5). [14] Summarizing (more details about this in Sec. 3.9), the Box trial move shows a *sharp* (first-order-like) transition at a value of $\sigma_{cr} \approx 1.75$, where an effective Arrhenius barrier $B_{eff}(\sigma)$ vanishes, and the gap starts to decrease as a power-law, $\Delta \propto T^{1/2}$.

One might suspect that the cusps shown in Fig. 3.3 for the Gaussian and Lorentzian cases signal, similarly to the Box case, some kind of transition. A closer inspection, however, shows that this is not so: it is just a level crossing phenomenon. In Fig. 3.6 we plot Δ versus the inverse temperature $1/T$ for several values of σ, for the Gaussian (Top panel) and Lorentzian (Bottom panel) proposal moves. After an initial Arrhenius-like behavior, particularly visible for the Gaussian small σ cases, the system always, and smoothly, changes to a low-T behavior which is entirely similar to the large-σ Box case: an apparent saturation of Δ followed by an avoided crossing between $\Delta = 1 - \bar{\lambda}_1$ and $1 - \bar{\lambda}_2$ (not shown) and a final $\Delta \propto T^{1/2}$ (see App. B.5). No real transition exist in the Gaussian and Lorentzian case: the cusp in Fig. 3.3 moves down toward smaller and smaller values of σ for decreasing T.

3.8 Results of Classical Annealing with optimal $\sigma(T)$

In the previous section we argued about the possibility of an optimal choice for the proposal move range $\sigma(T)$. This choice should guarantee the fastest *instantaneous* relaxation toward the instantaneous equilibrium distribution $P_e q(x)$ at any given temperature. The classical annealing performance is expected to be greatly improved by such a choice of $\sigma(T)$.

[14] Notice that at very large T the eigenvalues saturate. This is a consequence of the Metropolis algorithm: For $T \to \infty$, the acceptance factor $A(x', x)$ – which does not depends on the proposal move – goes to one, and therefore we are just diagonalizing the bare transition operator $T(x', x)$, which does not depend on T.

In practice, for each choice of proposal move (Box, Gaussian, Lorentzian), we have numerically determined the value of σ that maximizes the gap $\Delta = 1 - \bar{\lambda}_1$ for each fixed T. We will refer here to such an optimal σ as $\sigma_{opt}(T)$. In Fig. 3.7 we plot such an optimal choice for all the types of proposal moves introduced in Sec. 3.3. For comparison, we also show the $\sigma(T)$ we employed in Sec. 3.5, which is definitely smaller than σ_{opt}, for given value of T.

Having done this, we have then performed classical MC annealing runs, similarly to those reported in Sec. 3.5, with the usual linear annealing schedule for the temperature $T(t) = T_0\,(1 - t/\tau)$, where τ is the annealing time. In Fig. 3.8 we show the MC annealing results for the symmetric (Top) and asymmetric (Bottom) potential. The behavior for both choices of the potential is qualitatively similar, and we shall discuss them together. We stress again that the difference with respect to the runs illustrated in Sec. 3.5, where we took $\sigma = \sigma_0(T/T_0)^{1/2}$, is that, here, $\sigma_{opt}(T)$ is the optimal choice of σ obtained from the maximum instantaneous gap. In particular, the former choice was supposed to optimize the diffusion properties of the system, while the latter choice optimizes the instantaneous relaxation to the equilibrium distribution. For a theoretical link between these two approaches, see App. B.6. We notice that the Box and Gaussian results are now very different form the previous ones (see Fig. 3.2). Moreover, interestingly, Box and Gaussian data fall almost on the same curve, which is in turn very close to the Lorentzian case, the latter showing once again the best annealing results. [15]

3.9 A model for the low-lying spectrum of the transition operator

In this section we shall use analytical tools to explain the most relevant features of the spectra shown in Sec. 3.7.

[15] Notice that, in the Lorentzian case, the annealing exponent with $\sigma_{opt}(T)$ is actually a bit smaller than that obtained with the naive choice $\sigma \propto T^{1/2}$ (compare Figs. 3.2 and 3.8). This small discrepancy is the result of the MC implementation, and of a slight ergodicity loss at small T. Indeed, we have noticed that, for $\tau > 10^4$, the system always ends up, at the end of the Lorentzian annealing, in the lower minimum, without any component on the metastable minimum. Since, as it turns out, the Lorentzian $\sigma_{opt}(T)$ is slightly larger than the $\sigma_0\,T^{1/2}$ used in Sec. 3.5, this causes slightly larger fluctuations of the results around the lower minimum.

The question we want to address has to do with the low-lying spectrum of a Master equation of the form

$$P_{n+1}(x') = \sum_x W(x', x)\, P_n(x) \ , \tag{3.20}$$

where $W(x', x)$ is the transition operator associated to a certain choice of trial move in a Metropolis MC scheme (see Sec. 3.3). Equivalently, see the discussion in Sec. 3.1 and Eq. 3.8, we can study the spectrum of the exponential operator $e^{-\Delta t\, \bar{H}}$.

We shall start treating here the Box case, which admits a simple geometrical approach. For the general case, instead, we shall give here only the main steps of the procedure used to extract information on the gap of W, collecting the many technical details needed in the appendix of the present chapter.

3.9.1 The Box trial move: A simple geometrical interpretation

The Box trial move case shows an intriguing *sharp transition* of the gap at a critical value $\sigma_{cr} \approx 1.75$ (see Fig. 3.3). As shown in Fig. 3.9, the effective barrier $B_{eff}(\sigma)$ extracted from an Arrhenius fit, $\Delta = 1 - \bar{\lambda}_1(T) \propto e^{-B_{eff}(\sigma)/T}$, of the data presented in Sec. 3.7, goes to zero for $\sigma \to \sigma_{cr}$. There is a simple geometrical explanation for this critical behavior. First of all, we notice that the value of σ_{cr} is close to the distance between the two minima: this suggests that the transition is related to the availability of a *direct jump* from the bottom of the metastable well to the other one. If we call $x_\pm$, respectively, the minimum in the right and left well, there is a non-trivial solution of the equation $V(x_1) = V(x_+)$ with x_1 lying between the two minima, $x_- < x_1 < x_+$ (see Fig. 3.10): It turns out that $\sigma_{cr} = x_+ - x_1$.

Indeed, for $\sigma \geq \sigma_{cr}$ there is a possible proposed move that brings the system from $x = x_+$ (the metastable minimum) to the point $x' = x_1$ on the other side of the barrier: such a "non-local" move pays no energy ($\Delta E = 0$), and is therefore certainly accepted by the Metropolis algorithm. In some sense, for $\sigma > \sigma_{cr}$ the barrier is not seen (or, at least, there are allowed and accepted moves that do not see it), and B_{eff} is zero. Consider now the case $\sigma < \sigma_{cr}$. For every value $V = V(x_+) + \Delta V$, with $\Delta V > 0$, there are two equipotential solutions $x_1(\Delta V)$ and $x_2(\Delta V)$, such that $V(x_1) = V(x_2) = V$, lying between the two minima, but on opposite sides of the barrier, i.e., with $x_- < x_1(\Delta V) < x_2(\Delta V) < x_+$. Denote now by $d(\Delta V) = x_2(\Delta V) - x_1(\Delta V)$ the distance between such two equipotential points ($d(\Delta V)$ is

a monotonically decreasing function of ΔV). If the box width is σ, we can always find a ΔV such that $\sigma = d(\Delta V)$, i.e., such that there is a proposed move connecting $x = x_2$ to $x' = x_1$ which, once again, bypasses the barrier. The effective barrier seen by the system close to the metastable minimum, for a given value of σ, is therefore just (see Fig. 3.10)

$$B_{eff}(\sigma) = \Delta V = d^{-1}(\sigma) \; , \tag{3.21}$$

i.e., in essence, a piece of length σ is cut from the top of the barrier, and effectively not seen by the system. In other words, the effective barrier B_{eff} is just the potential energy drop ΔV the system must overcome before a long jump $|x' - x| \sim \sigma$ is made available at no energy cost ($V(x') \sim V(x)$). The value of $B_{eff}(\sigma)$ obtained through such a simple geometric construction is shown by a solid line in Fig. 3.9: the agreement with the numerical data is quite good, with small deviations for very small values of σ, which are likely due to the effect of the finite grid we employed in diagonalizing the transition operator W (see Sec. 3.7), and to finite T effects. [16]

We would like stress that this geometrical picture is strictly true only at zero temperature, and it does not apply to the Gaussian or Lorentzian cases, whose long tails provide a small non-vanishing chance of making a "long jump" for any value of σ. The treatment of the Gaussian and Lorentzian cases needs, therefore, a more detailed and technical discussion, to which we shall devote the following subsections.

3.9.2 General case: two-level system approximation

The general case of W associated to trial moves of the Gaussian or Lorentzian form is much harder to treat, and we have to resort to approximations. As explained in Sec. 3.1, a reduction to the exponential of a symmetric operator $e^{-\Delta t \bar{H}}$ – where the appropriate Δt is uniquely determined by the move range σ – is always possible.

The potential $V(x)$ will be always assumed to have the usual double-well form. At very small temperatures, the equilibrium distribution $P_{eq}(x)$ is concentrated around the local minima at $x = x_{\pm}$. Therefore, there is definitely a regime of parameters where a two-level system approximation for W must hold (see App. B.2). By exploiting this approximation, one can show (see appendix) that the spectral gap Δ of the transition operator W can be expressed in terms of a well-to-well propagator of

[16] For $\sigma \to 0$, we have that $B_{eff} \to V_0 - V(x_+) \simeq 0.9$, the geometrical barrier height.

the evolution operator of $\bar{H}$ as follows:

$$\Delta = \langle x_- | e^{-\Delta t \bar{H}} | x_+ \rangle \, e^{-\frac{V_- - V_+}{2T}} \ , \tag{3.22}$$

where T is the temperature and $V_\pm = V(x_\pm)$. The crucial quantity we need, therefore, is the well-to-well propagator $\langle x_- | e^{-\Delta t \bar{H}} | x_+ \rangle$. An expression for this propagator can be obtained by performing a Trotter decomposition and writing a Feynman integral. The final result can be cast in the form:

$$\langle x_- | e^{-\Delta t \bar{H}} | x_+ \rangle \propto \int_{V_+}^{V_0} \mathrm{d}\, V \, \frac{\partial\, d}{\partial\, V} \, K_1(d(V - V_+); \Delta\, t) \, e^{-\frac{1}{T}\left(V - \frac{V_+ + V_-}{2}\right)} \, T(d(V - V_+)) \ , \tag{3.23}$$

where $d(V - V_+)$ is the function introduced in Sec. 3.9.1 (see Fig. 3.10), and K_1 accounts for the Gaussian fluctuations around the saddle-point classical solution of the Feynman integral, which do not modify the important exponentially activated behavior provided by the term $e^{-\frac{1}{T}\left(V - \frac{V_+ + V_-}{2}\right)}$. $T(d(V - V_+))$, finally, is just the proposal move distribution: $T(d(V - V_+)) = T(|x_2 - x_1|) = T(x_2, x_1)$ (see Sec.3.3).

This rather simple equation for the well-to-well propagator can be further treated by a standard saddle-point approximation of the relevant integral. We shall discuss here the cases we are interested in: Box, Gaussian, and Lorentzian.

Box move: T was defined in Eq. 3.15. We recall that the function $d(V - V_+)$, introduced in Sec. 3.9.1, is a monotonically decreasing function of its argument, which exhibits an infinite first order derivative for $V = V_0$. There is no true saddle-point for such a move, since the functions involved in the exponential part of Eq. B.8 are all monotonic. On the other hand, – in the case of the Box move – the extremes of integration in Eq. 3.23 can be taken as $[V_+, d^{-1}(\sigma) + V_+]$ instead of $[V_+, V_0]$, since its transition matrix, $T(d(V - V_+))$, is zero outside the former interval. Therefore, the largest contribution to the whole integral is due either to $V = V_0$ or $V = d^{-1}(\sigma) + V_+$, which are the two integration extremes. It turns out that the latter choice guarantees the largest exponential, and is therefore the right one. As a consequence, we can write down that:

$$\langle x_- | e^{-\Delta t \bar{H}} | x_+ \rangle \propto e^{-\frac{1}{T}\left(d^{-1}(\sigma) + \frac{V_+ - V_-}{2}\right)} \ .$$

This transition amplitude is Arrhenius-like, and so does the gap function (see Eq. B.5) which reads:

$$\Delta(T, \sigma) \propto e^{-\frac{B_{eff}(\sigma)}{T}} \ ,$$

where $B_{eff}(\sigma) = d^{-1}(\sigma)$, as anticipated in Sec. 3.9.1 from geometrical considerations. This is indeed the behavior seen in Fig. 3.4.

Gaussian move: T was defined in Eq. 3.16. In this case, a true saddle-point is possible if the following equation has a solution:

$$\frac{1}{T} - \frac{d(V - V_+)}{\sigma^2} \frac{\partial}{\partial V} d(V - V_+) = 0 \, . \tag{3.24}$$

Since $\frac{\partial d}{\partial V} < 0$ and $d > 0$, then, in the limit $T \to 0$ we have that either $V = V_0$ or $V = V_+$ are solutions of Eq. 3.24. The second choice guarantees the largest contribution to the integral in Eq. B.8, and must be therefore selected. The resulting expression for the transition amplitude is:

$$\langle x_- | e^{-\Delta t \, \bar{H}} | x_+ \rangle \propto e^{-\frac{V_+ - V_-}{2T}} \, e^{-\frac{d(0)^2}{2\sigma^2}} \, ,$$

and the corresponding gap value is:

$$\Delta(T, \sigma) \propto e^{-\frac{d(0)^2}{2\sigma^2}} \, . \tag{3.25}$$

Fig. 3.11 shows that such a dependence perfectly fits the simulation data. We emphasize that even the coefficient within the exponential part of the fitting function matches the theory (we recall that $d(0) \simeq 1.75$): only an overall prefactor is used as fitting parameter.

As observed in Sec. 3.7, the previous equation will be no longer valid for very small values of the temperature T, where an (avoided) crossing with the eigenvalue λ_2 will occur. On the other hand, for large values of σ we see from Fig. 3.11 that the data deviate from the behavior predicted by Eq. 3.25. This is the consequence of another level crossing (see App. B.5).

Lorentzian move: T was defined in Eq. 3.17. To cut a long story short (calculations proceed similarly to the Box and Gaussian case), we write down the result. The gap value is:

$$\Delta(T, \sigma) \propto \sigma \, , \tag{3.26}$$

and also in this case the model agrees with the data, see Fig. 3.11, within a multiplicative fitting coefficient.

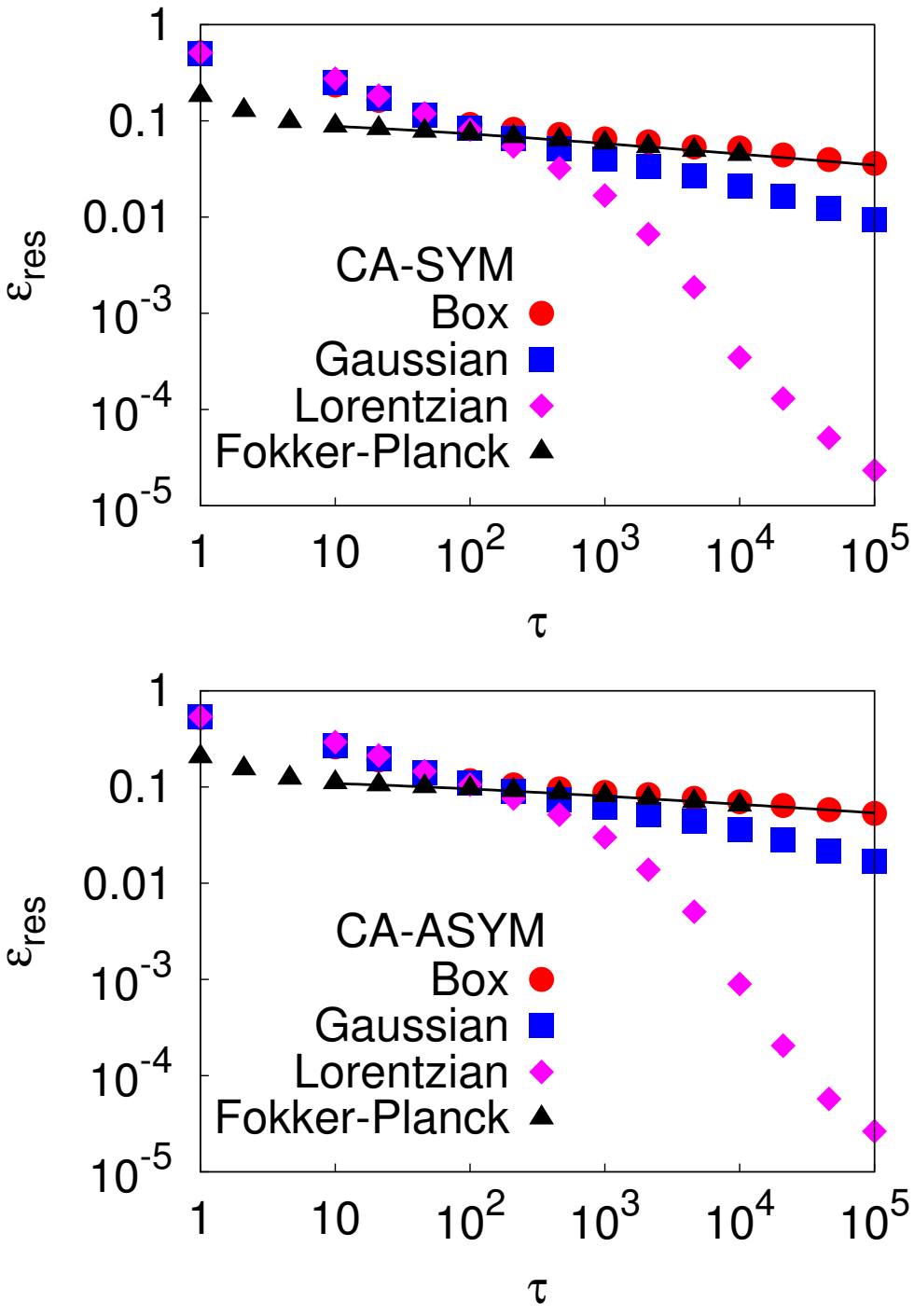

Figure 3.2: Plot of a Monte Carlo implementation of the classical annealing of a double-well potential, using three types of proposal moves: Box, Gaussian, and Lorentzian. τ is the annealing time and ϵ_{res} the residual energy at the end of the annealing (see text). We also report the results of the Fokker-Planck dynamics, for comparison. These data refer to the symmetric (SYM, Top) and to the asymmetric (ASYM, Bottom) double-well potential.

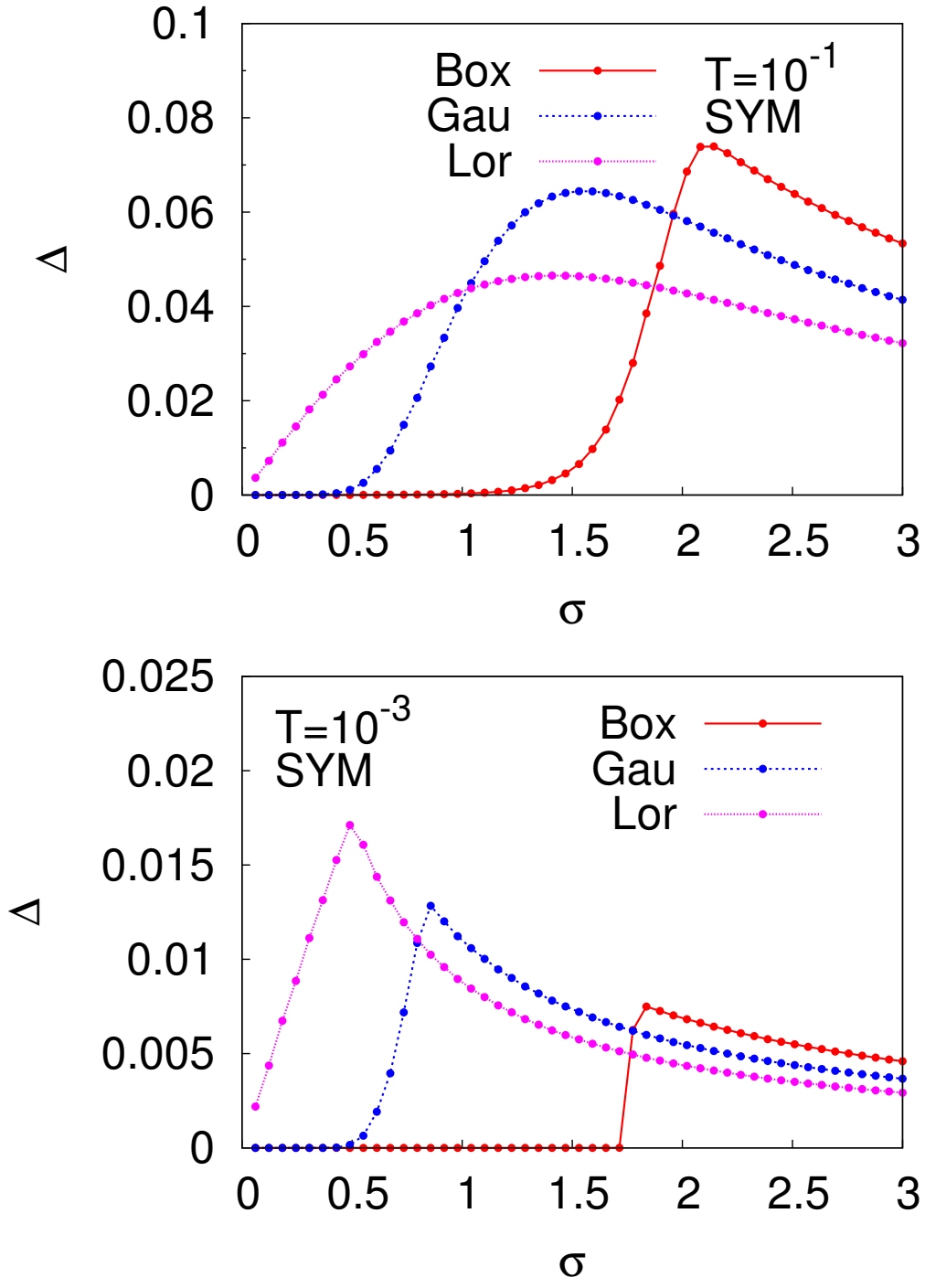

Figure 3.3: Exact diagonalization of the transition operator W, for the symmetric potential case. We plot the gap $\Delta = 1 - \bar{\lambda}_1$ versus the proposal move range σ. The temperature is $T = 0.1$ (Top) or $T = 0.001$ (Bottom), in the same units in which the barrier parameter is $V_0 = 1$ (see Sec. 2.2).

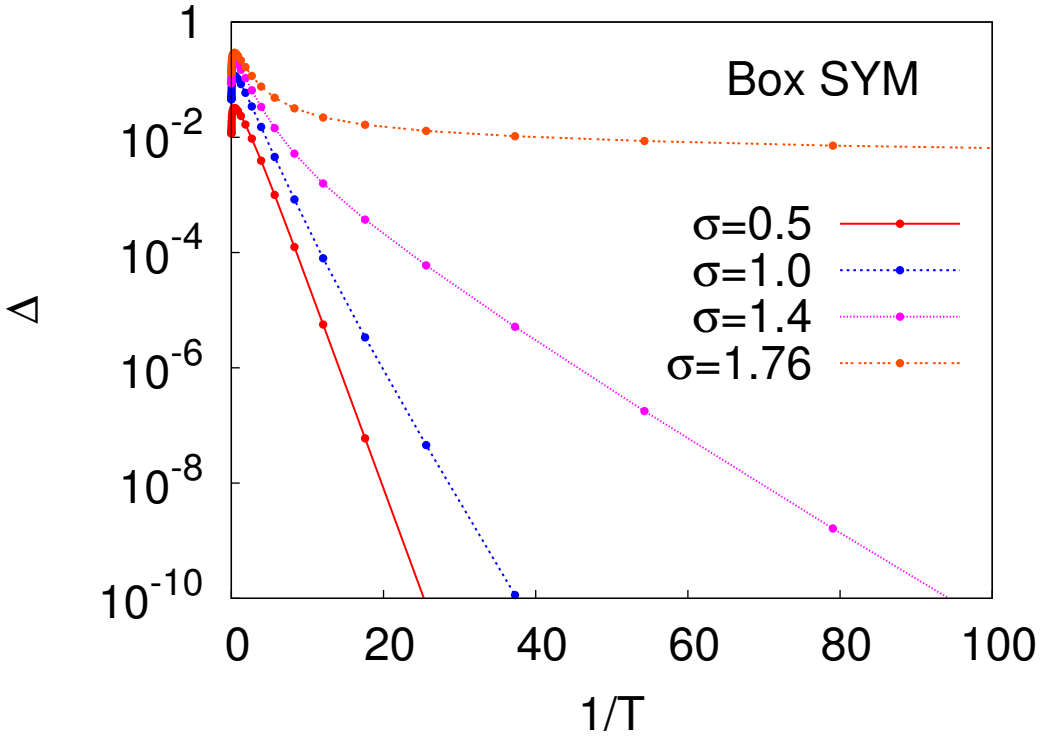

Figure 3.4: Exact diagonalization of the transition operator W, for the symmetric potential, with Box proposal move. We plot the gap $\Delta = 1 - \bar{\lambda}_1$ versus the inverse temperature $1/T$ for several fixed values of the proposal move range σ.

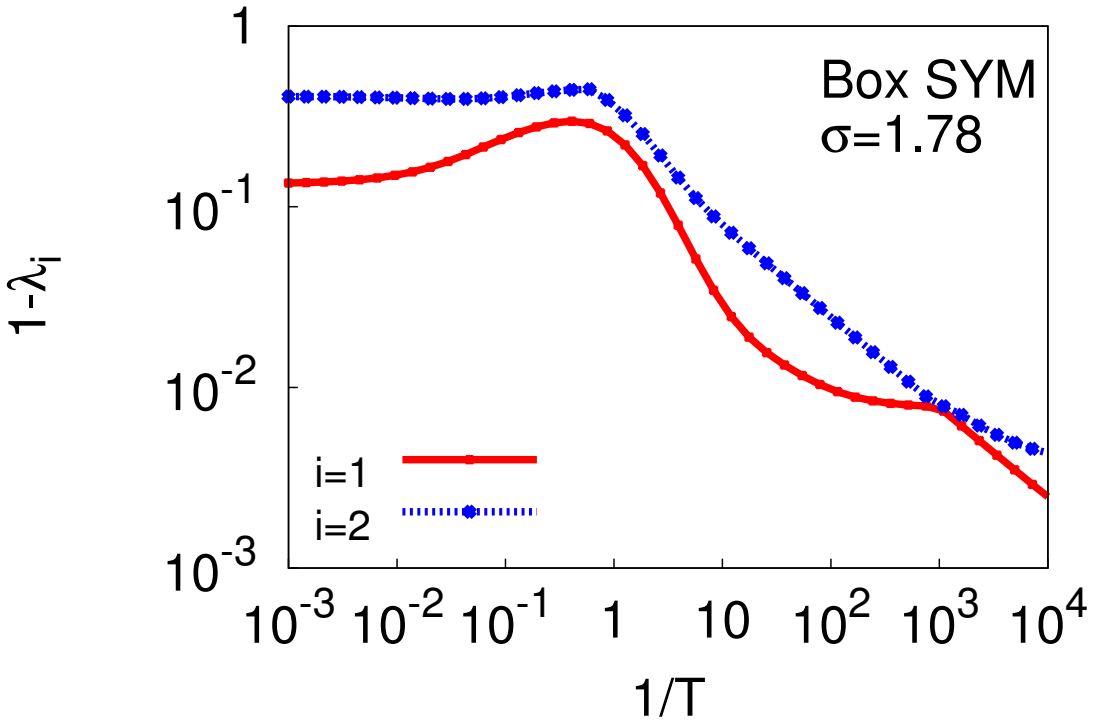

Figure 3.5: Exact diagonalization of the transition operator W, for the symmetric potential, with Box move. We plot the gap $\Delta - 1 - \bar{\lambda}_1$ and the next eigenvalue $1 - \bar{\lambda}_2$ versus the inverse temperature $1/T$ for $\sigma = 1.78 > \sigma_{cr}$.

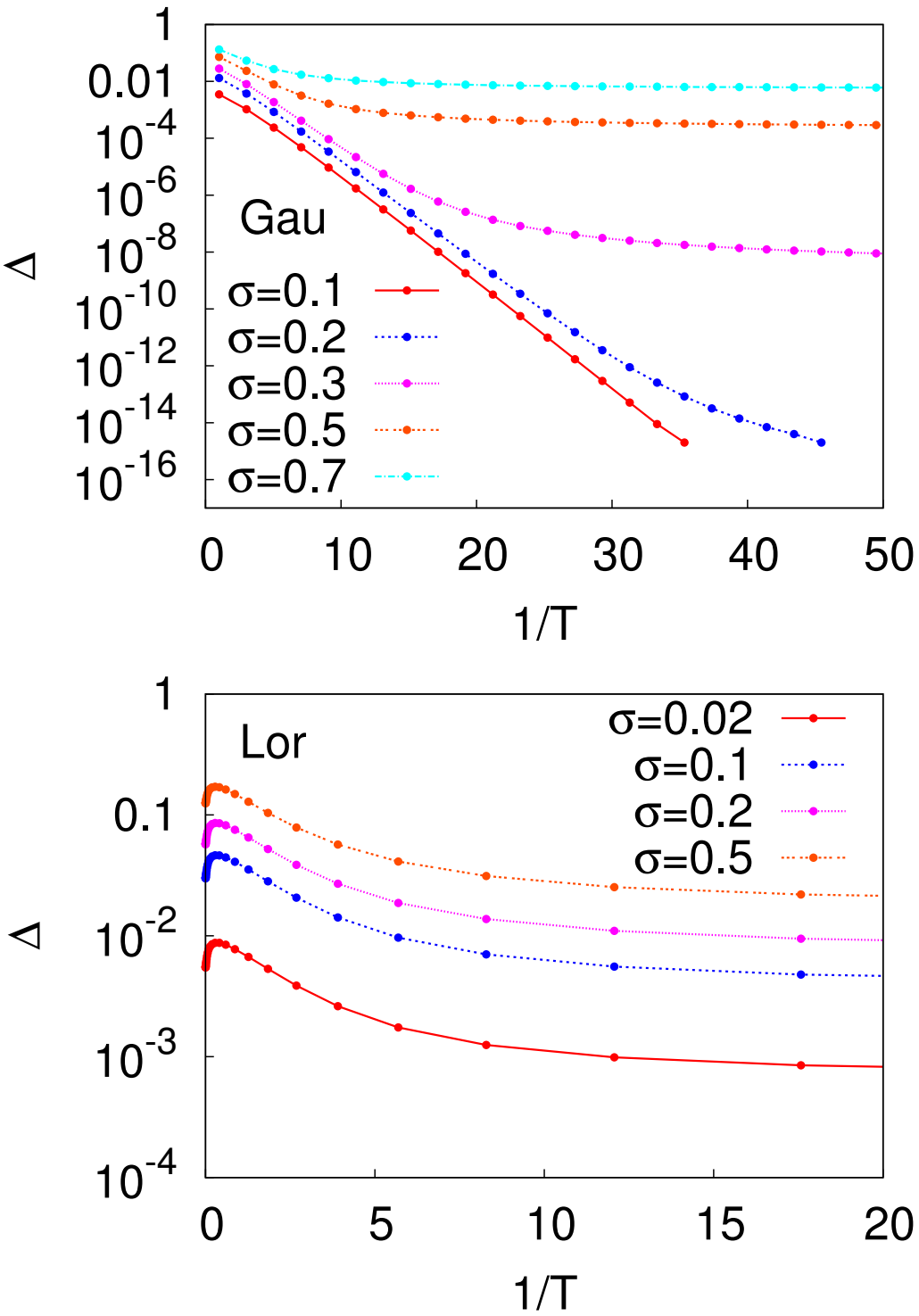

Figure 3.6: Exact diagonalization of the transition operator, for the symmetric potential, with Gaussian (Top) or Lorentzian (Bottom) move. We plot the gap Δ versus the inverse temperature $1/T$ for several fixed values of the proposal move range σ.

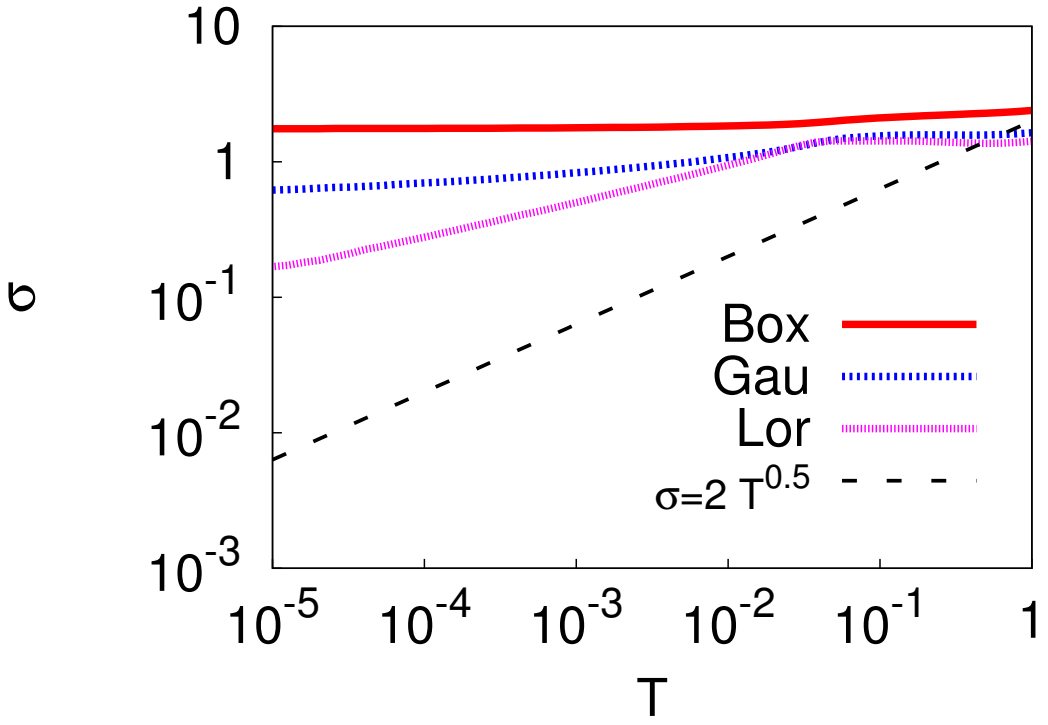

Figure 3.7: Plot of the optimal $\sigma_{opt}(T)$ for the Box, Gaussian and Lorentzian proposal move. We also show, for comparison, the schedule $\sigma(T) \propto T^{1/2}$ employed in Sec. 3.5.

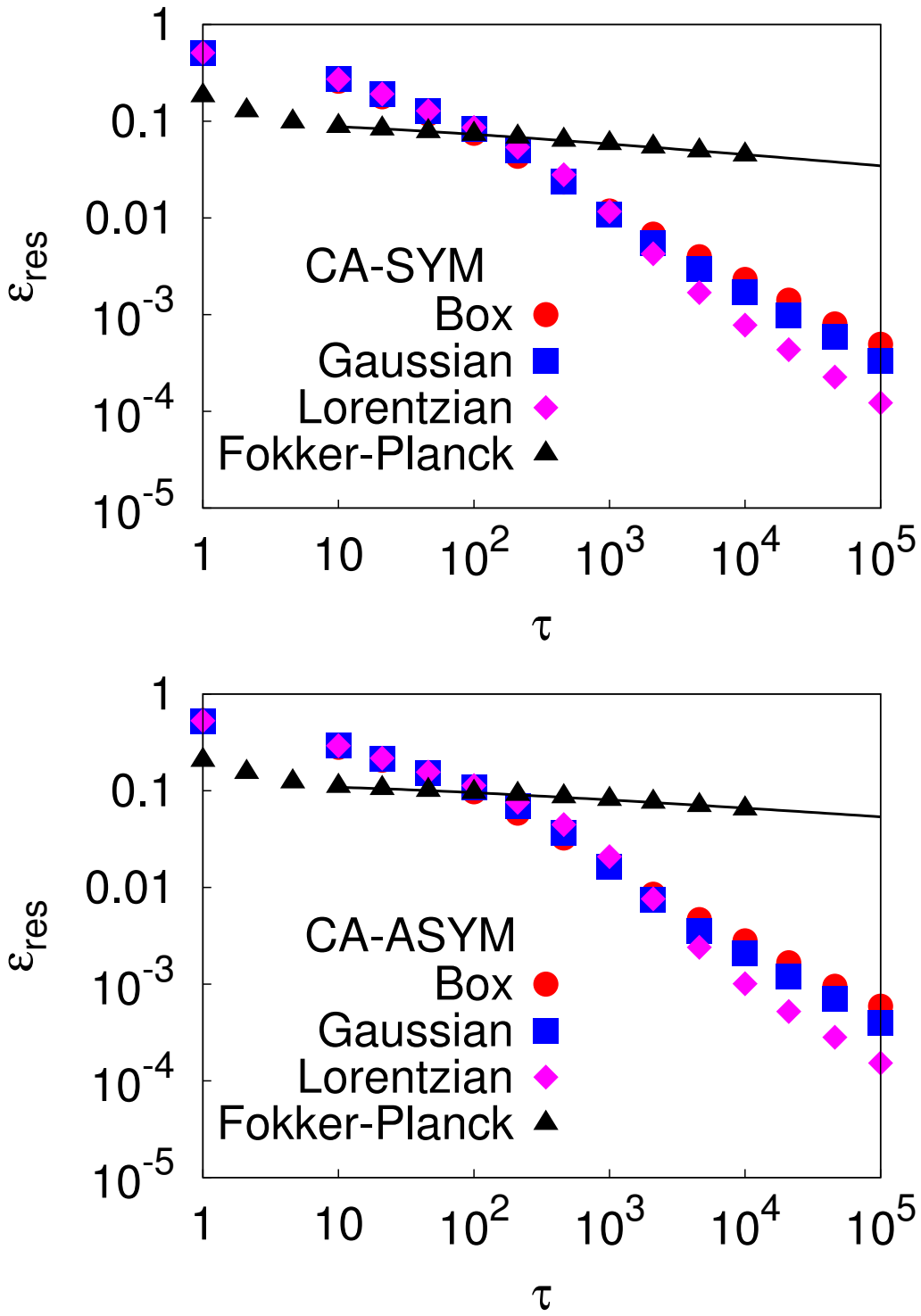

Figure 3.8: Plot of Monte Carlo classical annealings for the symmetric (Top) and asymmetric (Bottom) potential. τ is the annealing time and ϵ_{res} the residual energy (see text). We report the results of the exact integration (Fokker-Planck), together with the actual MC data for several proposal moves (Box, Gaussian, Lorentzian). The MC simulations are performed with an optimal choice for $\sigma(T)$ obtained from the maximum instantaneous gap (see text).

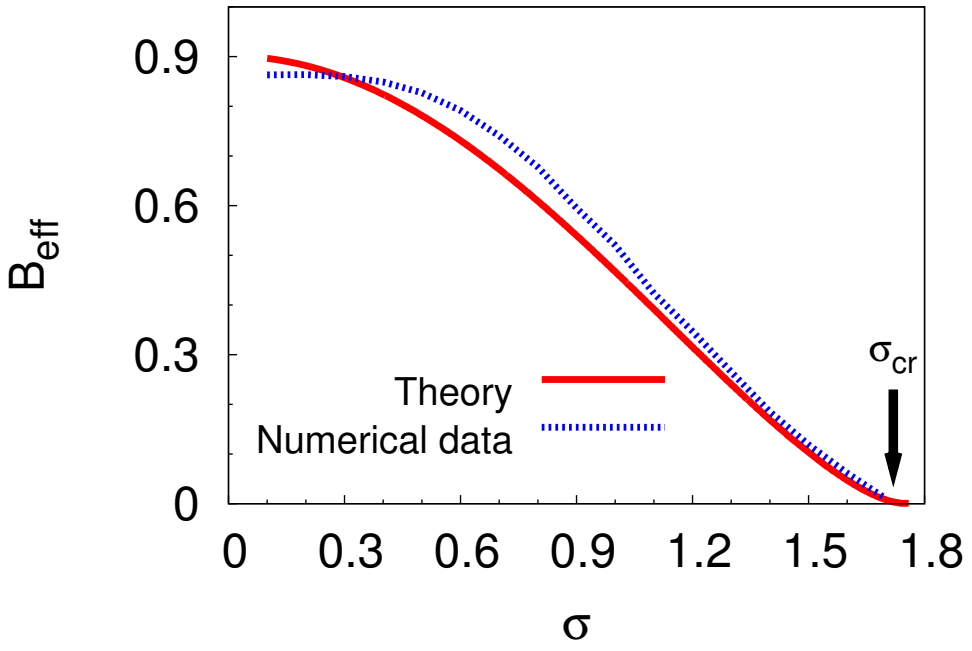

Figure 3.9: Plot of the effective barrier $B_{eff}(\sigma)$ versus σ for the Box case. The data are extracted from an Arrhenius fit of the diagonalization data presented in Sec. 3.7, while, for the theory, $B_{eff}(\sigma) = d^{-1}(\sigma)$, see the text. An arrow indicates the position of the critical value σ_{cr} such that $B_{eff}(\sigma_{cr}) = 0$.

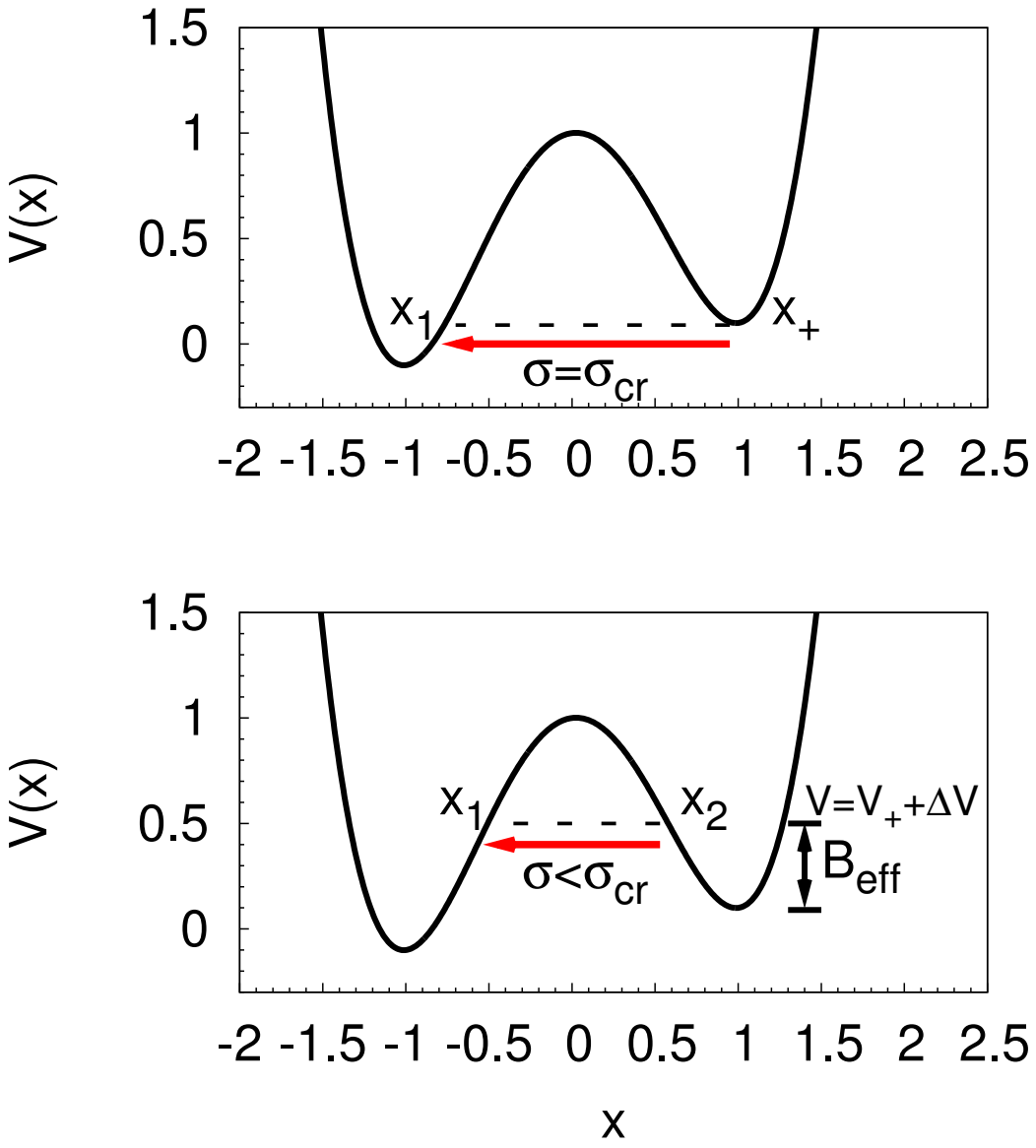

Figure 3.10: Illustrative sketch of a Box move in a double-well potential. In the upper panel we consider the case $\sigma = \sigma_{cr}$, while in the lower one we illustrate the general case $\sigma < \sigma_{cr}$. Here $V_+ = V(x_+)$ is the potential at the bottom of the metastable minimum. The meaning of the other symbols is explained in the text.

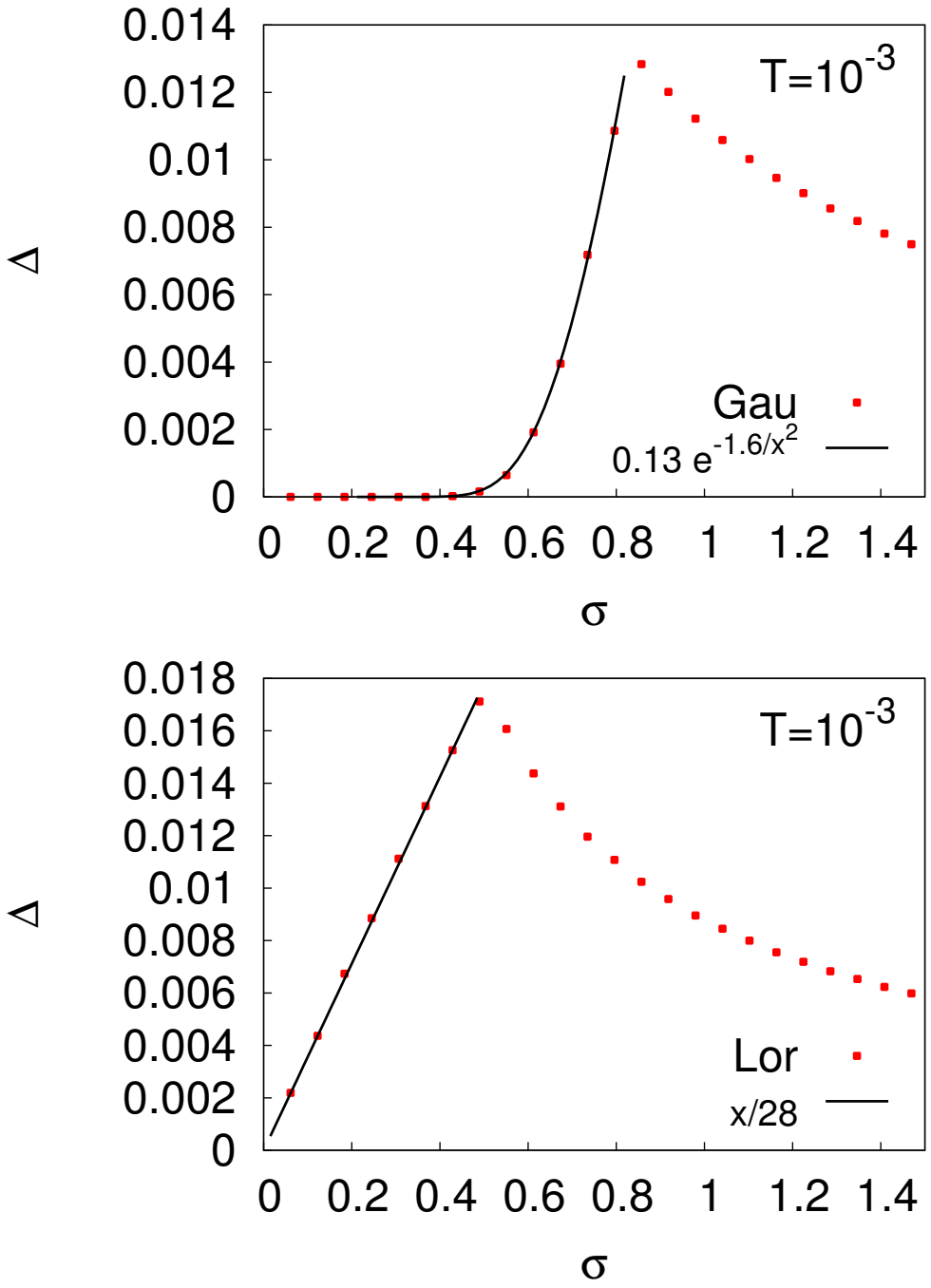

Figure 3.11: Plot of the gap function Δ for the Gaussian (Top) and Lorentzian (Bottom) proposal move versus the move range σ obtained by numerical diagonalization (see Sec.3.7). The dashed line represents the fit provided by Eq. 3.25 (Top) and Eq. 3.26 (Bottom), the only fitting parameter being an overall multiplicative constant.

Chapter 4

Annealing by Path-Integral Monte Carlo

In Chap. 2 we studied, by direct numerical integration, the (real- and imaginary-time) Schrödinger dynamics provided by the *time-dependent* Hamiltonian

$$H = -\Gamma(t)\,\frac{\partial^2}{\partial\, x^2} + V(x)\ , \tag{4.1}$$

where $\Gamma(t) = \hbar^2/(2m(t)) = \Gamma_0\,\left(1 - \frac{t}{\tau}\right)$ is the inverse mass parameter, providing the strength of the quantum fluctuations. This is, in some sense, a paradigmatic implementation of a quantum annealing (QA) strategy. As pointed out in many circumstances, this Schrödinger dynamics is a kind of *Gedanken* application of QA, suitable only for toy problems with a very limited Hilbert space, and inapplicable on actual hard optimization problems. In order to become a viable strategy for actual optimization problems, QA has to resort to *stochastic*, i.e., Quantum Monte Carlo (QMC), approaches, mostly appropriate for an *imaginary-time* framework (we remind the reader that, as shown in Chap. 2, working in imaginary time is actually beneficial for QA).

There are several possible QMC techniques on which a QA strategy can be build. By far the simplest of these QMC techniques is the *Path-Integral Monte Carlo* (PIMC) method, which has already been used with some success in the QA context [10, 11, 13, 16]. The method does not addresses the imaginary-time Schrödinger time-dependent dynamics, but simulates an *equilibrium quantum system*, held at a small finite temperature T, where the relevant parameter $\Gamma(t)$ is externally turned off in the course of a QMC simulation: t is not treated, therefore, as a proper physical time, but only as a

Monte Carlo time. [1]

In the present chapter we will explore the potential of a PIMC-based QA strategy on the same simple toy problem treated so far: the one-dimensional double well potential. The reason for investing effort on such a simple problem, is that the simplicity of the problem landscape will allow us to have a perfect control of all the ingredients of the method. Indeed, we will learn a lot about PIMC-QA, particularly the limitations of the method, from investigating such a simple toy problem.

We give a summary of the results and a few final comments in Sec. 4.5.

4.1 The Path-Integral Monte Carlo

The equilibrium properties of our quantum system, for any fixed value of $\Gamma = \Gamma(t)$, are all encoded in the quantum partition function

$$Z(\beta) = \mathrm{Tr}\left[e^{-\beta\,(\hat{T}+\hat{V})}\right] = \int dx\ \langle x|e^{-\beta(\hat{T}+\hat{V})}|x\rangle \quad \text{where} \quad \hat{T} = -\Gamma\,\frac{\partial^2}{\partial\, x^2}\ , \tag{4.2}$$

from which all the thermodynamics follows, for instance the internal energy being given by:

$$U(\beta) = -\frac{\partial}{\partial\,\beta}\,\ln Z(\beta)\ . \tag{4.3}$$

Here, as usual, $\beta = (k_B\,T)^{-1}$. We will drop the Boltzmann constant k_B from now on. As we see from Eq. 4.2, Z involves an integral over a positive distribution, $\langle x|e^{-\beta(\hat{T}+\hat{V})}|x\rangle$, which involves, however, the very difficult task of calculating the diagonal matrix element of the *exponential* of the Hamiltonian H. We will now show that the calculation of the exponential of the Hamiltonian reduces, at the price of an enlarged configuration space, to a classical partition function, for which all the standard tools of classical MC apply.

The standard approach is to rewrite Eq. 4.2 as a Feynman's *Path-Integral* [58]. The main mathematical tool employed in such an approach is the so-called *Trotter theorem*, which reads:

$$e^{-\beta(\hat{T}+\hat{V})} = \lim_{P\to\infty}\left(e^{-\frac{\beta}{P}\hat{T}}\,e^{-\frac{\beta}{P}\hat{V}}\right)^P\ , \tag{4.4}$$

[1] This strategy parallels in a direct way the philosophy behind the classical Simulated Annealing technique, where the temperature $T(t)$ is also a parameter which is externally driven in the course of a standard equilibrium MC simulation.

where P is the number of *Trotter's slices*, in which the imaginary time interval $[0, \hbar\beta]$ has been partitioned, each slice being of length $\Delta t = \frac{\hbar\beta}{P}$. By means of Eq. 4.4, and inserting $P-1$ identities in the form of $\mathbf{1}_i = \int dx_i |x_i\rangle\langle x_i|$, one can rewrite the partition function (see App. C.1 for this simple textbook derivation) as:

$$Z(\beta) = \left(\frac{1}{4\,\pi\,\Gamma\,\Delta\,t}\right)^{\frac{P}{2}} \int \prod_{i=0}^{P-1} \mathrm{d}x_i\, e^{-S_{PA}[x]} + O\left(\Delta\,t^3\,P\right) \ , \tag{4.5}$$

where $S_{PA}[x]$ is the so-called (euclidean) *primitive action*:

$$S_{PA}[x] = \Delta\,t \sum_{i=0}^{P-1} \left\{ \frac{m}{2} \left(\frac{x_{i+1} - x_i}{\Delta\,t}\right)^2 + V(x_i) \right\} \quad \text{with} \quad x_P = x_0 \ . \tag{4.6}$$

(Notice the periodic boundary condition involved in Eq. 4.6, $x_P = x_0$, a consequence of the trace present in Eq. 4.2.)

The similarity of Eq. 4.5 with the classical partition function of a kind of *closed polymer* with P beads is evident: the polymer beads x_i can be seen as the imaginary-time snapshots $x_i = x(i\Delta\,t)$, $i = 1, \cdots, P$, of a fluctuating closed path $x(t)$ in the enlarged configuration space (x, t), where t is the imaginary time. Two neighboring beads x_i and x_{i+1} interact with harmonic interactions of spring constant $K^{\perp} = mP^2/(\hbar\beta)^2$, originating from the propagator of the quantum kinetic term,

$$\langle x_i | e^{-\frac{\beta}{P}\hat{T}} | x_{i+1} \rangle \propto e^{-\frac{\beta}{P}\frac{K^{\perp}}{2}(x_i - x_{i+1})^2} \ . \tag{4.7}$$

The strength of the harmonic interactions in imaginary time controls the amount of quantum fluctuations: a large mass m (classical regime) results in a strong $K^{\perp}$ and hence in a very "rigid" polymer, while a small mass m (quantum regime) results in a very soft and fluctuating polymer. All beads are also subject to the classical potential $V(x)$. Once we have reduced our problem to the Path-Integral form in Eq. 4.5, the standard techniques of classical Metropolis MC can be used, and the resulting algorithm is what is called a *Path-Integral Monte Carlo* (PIMC). The most obvious MC moves to be used are just *single bead* moves, exactly as in a classical MC (see Chap. 3).

These are the bare-bones of a PIMC approach. A similar derivation applies to the Ising problem in a transverse field, with the difference that the spring constant $K^{\perp}$ turns into a ferromagnetic interaction, while single bead moves are replaced by single spin-flip moves. For the problem we are dealing with, a particle in a potential (or

more generally for systems of quantum particles on the continuum), one can improve the method just sketched in two possible directions: i) by improving the quality of the approximation in Eq. 4.5, so as to get a smaller Trotter discretization error for a given number of Trotter slices P used; ii) by introducing MC moves which are more sophisticated than just moving a single bead at a time. Regarding i), we will adopt a *fourth order* approximation to the action, which improves the Trotter truncation error of the partition function from $O\left(\Delta\, t^3\, P\right)$ to $O\left(\Delta\, t^5\, P\right)$. This leads to an improvement of the internal energy error from $O\left(\Delta\, t^2\right)$ to $O\left(\Delta\, t^4\right)$. As for the MC move choice (ii), we will adopt smarter moves, known in the context of classical polymer simulations, which reconstruct entire pieces of the polymer, instead of a single bead at a time. We will also appreciate the importance of *instanton moves*.

We now move to the presentation of the results.

4.2 Results I. IV-order action and bisection move

The error done in calculating physical quantities, like the internal energy, by employing the primitive approximation – Eqs. 4.5 and 4.6 – is proportional to $O\left(\Delta\, t^2\right)$, where $\Delta t = \hbar\beta/P$. This error can be easily improved: many algorithms which achieve *fourth-order* precision in $\Delta\, t$, $O\left(\Delta\, t^4\right)$, have been devised. [2] We refer to App. C.2 for details on this topic. We have adopted the so-called *Takahashi-Imada* approximation, which is especially suitable for continuum systems. The partition function is still given by an expression of the form Eq. 4.5 with the primitive action S_{PA} replaced by the following Takahashi-Imada action:

$$S_{TIA}[x] = \Delta\, t \sum_{i=0}^{P-1} \left\{ \frac{m}{2} \left(\frac{x_{i+1} - x_i}{\Delta\, t} \right)^2 + V_{eff}(x_i) \right\} \quad \text{with} \quad x_P = x_0 \ , \tag{4.8}$$

where the only difference with respect to S_{PA} is in the potential energy, which now reads:

$$V_{eff}(x) = V(x) + \frac{1}{12} \Gamma \left(\Delta\, t\right)^2 \left(\frac{\partial V(x)}{\partial\, x} \right)^2 \ . \tag{4.9}$$

[2] Since we want to use the smallest possible T (large β), and the computation cost of the PIMC algorithm scales with P, the use of a fourth-order approximation leads to a sensible speed-up of the QA simulations.

Concerning, briefly, the other details of the implementation, we have employed tools developed in different physical contexts [3] in order to obtain the best performing implementation of a PIMC-QA. In particular, we made use of the *virial-centroid estimators* of the statistical functions (see App. C.3) in order to minimize the fluctuations in the averages. Moreover, we took advantage of the so-called *bisection algorithm* (see App. C.6) in order to propose a more effective set of moves involving many Trotter's replicas. [4] This choice guarantees a fast relaxation to the instantaneous equilibrium distribution. Finally, we also used *global (classical) moves*, in the form of rigid motion of the center of mass of the polymer (See App. C.7), so as to keep a good sampling in the final part of the annealing, where the mass m is large (Γ is small) and quantum moves are suppressed.

Several tests were performed in order to set up the parameters that govern the accuracy and the efficiency of the algorithm. We refer the reader to App. C.8, Secs. C.9 and C.10, for a more detailed discussion of these issues. As for the inverse-mass annealing parameter $\Gamma(t) = \frac{\hbar^2}{2\,m(t)}$ appearing in Eq. 4.8, we decrease it linearly to zero, $\Gamma(t) = \Gamma_0 \left(1 - \frac{t}{\tau}\right)$ (it is understood here that both t and the total annealing time τ are MC times, i.e., measured in units of MC steps, each MC step being made of one bisection move plus one global move). The initial condition is set to $\Gamma_0 = 0.5$ (as in Sec.2.2). For every value of the annealing time τ we calculated the relevant averages by repeating several times the same annealing experiment, starting from different randomly distributed initial conditions (see App.C.10).

In Fig. 4.1 we plot the final PIMC-QA residual energy obtained for both potential choices, V_{asym} (Top) and V_{sym} (Bottom) (with the same parameters used in Chap. 2), for a fixed temperature $T = 0.03$. The statistical errors are evaluated by 10^3 repetitions of each annealing run for $\tau < 10^6$, and 10^2 repetitions for $\tau > 10^6$. We employed $P = 160$ with an $l = 5$ bisection level (i.e., moving polymer pieces containing $2^5 - 1$ beads, see App. C.6) for $\tau \leq 10^5$, and $P = 20$ with $l = 2$ (moving $2^2 - 1$ beads) for $\tau > 10^5$. We stress that the two series of data match perfectly, but obviously the use of a smaller P allows us to achieve larger annealing times τ, which

[3] Mainly in the simulation of quantum fluids. See Ref. [59] for a review of the subject.

[4] Actually, another sampling technique, the *staging* algorithm [60], has been previously employed in the work of Liu and Berne [9] on Quantum Annealing of Lennard-Jones clusters. The staging method is based on ideas very similar to those of the bisection method and it is also belied to be equally effective [59][page 331], although it requires a bit more involved numerical computations.

would be otherwise too computationally heavy. (An analysis of the convergence behavior of the algorithm as a function of both Γ and P can be found in App. C.8.) In the V_{asym} case (Top panel of Fig. 4.1) – which, we recall, presents a Landau-Zener (LZ) transition (see Sec. 2.2.1) –, the residual energy $\epsilon_{res}(\tau)$ as a function of the annealing time τ gets stuck around $\epsilon_{res} = 0.2$. Since this energy is comparable to the energy of the metastable state, $V(x_+)$, and definitely larger than the thermal limit $T/2 = 0.015$, we can claim that the algorithm failed to follow adiabatically the ground state. For comparison, we also reported in Fig. 4.1 the residual energy data obtained by the exact imaginary-time Schrödinger annealing (see Sec. 2.2), even if the time-scales of the two algorithms are definitely different and unrelated. Considering, on the other hand, the symmetric potential case V_{sym}, bottom panel of Fig. 4.1, we notice that the algorithm succeeds in reaching the thermal limit $T/2$ in a reasonable amount of MC steps.

From this first comparison between PIMC-QA and exact Schrödinger QA, we appreciate that the LZ transition leads to a severe slowing down of the Monte Carlo algorithm, the actual tunneling event being essentially missed by the PIMC algorithm.

This difficulty is also present in a *static* simulation performed in the neighborhood of the LZ crossing (occurring at $\Gamma \sim 0.038$), where we find (see App. C.9) a dangerous loss of adiabaticity which calls for an improved sampling of the action: we will show how this improved sampling is achieved by the introduction of *instanton moves*.

4.3 Results II. Adding the instanton move

In the previous section we tested an allegedly *state-of-the-art* PIMC-QA algorithm for the very simple problem of a particle in a double-well potential, obtaining disappointingly bad results. The essential problem is a *sampling crisis*: our action is accurate, but its sampling does not accurately accounts for the tunneling events, and this is catastrophic when a LZ crossing occurs (asymmetric potential case). The well-known cure for such a sampling problem, for the case of a *perfectly symmetric* double-well, $V(x) = V_0(x^2 - a^2)$, is the introduction of the so-called *instanton move* (see App. C.11). In a nutshell, an instanton is defined has a solution of the classical equation of motion in the inverted potential which goes from a minimum to the other. Moreover, the time-reversed path (anti-instanton) is also a solution. Strictly speaking, the whole classical trajectory takes an infinite (real) time, while the barrier

overcoming is a very fast process (whence the name instanton). In our implementation, we made use of the imaginary-time version of an instanton/anti-instanton pair as a proposal MC move (see App. C.11). In practice, the instanton move proposes to a subset of the Trotter's slices an excursion from one minimum to the other. The move will be accepted according to the usual Metropolis criterion, with an energy competition between the possible gain in potential energy compensated by the increase in kinetic energy due to the spring stretchings (at the positions of the instanton and of the anti-instanton). [5] Obviously, the instanton move will be not effective in the classical limit (small Γ, final part of the annealing), since it describes an inherently quantum effect: the tunneling precess between the two wells. The interested reader is referred to App. C.11 for more insights.

Before showing the results obtained, we need to stress an important point: the instanton move is available only for potentials which are small deformations of a perfectly symmetric double-well potential – as $V_{sym}(x)$ and V_{asym} are, see Sec. 2.2– and is not the general key for solving sampling problems (ergodicity breaking) for a generic potential, whose landscape is generally poorly known. In essence, we are playing here an unfair game: we are using a *detailed information* on the potential landscape in order to correctly implement the tunneling dynamics in our PIMC.

Fig. 4.2 shows the PIMC-QA residual energy results when the instanton move is introduced, again for the case of V_{asym} (Top) and V_{sym} (Bottom). As in the previous section, we employed $P = 160$ Trotter's slices with $l = 5$ bisection level, for $\tau \leq 10^5$, and $P = 20$ with $l = 2$ bisection level for $\tau > 10^5$. Statistical errors are evaluated with 10^3 repetitions of every annealing run. We notice that the introduction of the instanton move causes a sensible improvement of the residual energy slope for the V_{asym} potential case: the asymptotic power-law behavior is now quite evident, $\epsilon_{res}(\tau) \propto \tau^{-\Omega_{QA}}$, although with an exponent which is only $\Omega_{QA} = 0.19$ (smaller than in the Schrödinger case). For the symmetric potential case, the instanton move leads to a faster convergence to the thermal limit, but does not give a big qualitative change.

We want to stress that the partial acceptance ratio of the instanton moves is very low – around 1% –, despite the fact that this proposal move was quite "tailored" on the

[5] Even if the harmonic stretching is shared by more that two single bonds, the kinetic energy cost may be large whenever a too small number of Trotter slices, P, is employed. Nevertheless, the potential energy can compensate for this loss, provided the $\Gamma = \frac{\hbar^2}{2\,m}$ is not too small.

potential under investigation. [6] Nevertheless, such a move changes quantitatively the PIMC-QA performance. This is likely due to a somewhat generic feature of systems with barriers: the importance of rare events in the long-time structural properties.

Unfortunately, the instanton "recipe" – which is natural for a double-well potential – cannot be generalized to generic landscapes, including more complicated potentials, not to mention actual combinatorial optimization problem. Nevertheless, the results obtained are instructive in many respects: they show clearly a big potential limitation of the PIMC approach, while demonstrating, once more, that a smart choice of the proposal move – even in the case of PIMC-QA – leads to a sensible improvement of the annealing performance.

4.4 Results III. Changing kinetic energy: The Lorentzian Move

It is by now a sort of *Leitmotiv* of the previous and current chapter, that the choice of the MC move strongly influences the dynamics, and hence the annealing behavior. We recall, in particular, see Chap. 3, that in the classical annealing case the winning choice was given by a Lorentzian distributed proposal move, which seldom provides very long displacements.

In the present PIMC context, however, the non-relativistic kinetic part of the Hamiltonian Eq. 4.1 *forces*, in a sense, the choice of Gaussian distributed moves, because the free propagator of the non-relativistic kinetic energy is a Gaussian, see Eq. 4.7. [7] The whole bisection algorithm makes strong use – see App. C.4, Secs.C.5 and C.6 – of the Gaussian nature of the free propagator.

It is natural to ask what would be the QA behavior if, instead of the standard non-relativistic kinetic energy used so far, we employ a *relativistic* Hamiltonian of the form

$$H = \Gamma(t)\,|p| + V(x)\ . \tag{4.10}$$

[6] A possible explanation is the inherent lack of global symmetry of our potentials with respect to the perfectly symmetric double-well case, where the instanton and anti-instantons trajectory are well-defined mathematical objects (see App. C.11).

[7] This is, as previously noticed, the origin of the harmonic coupling between neighboring Trotter replicas.

One immediate consequence of this choice is that the free Gaussian propagator appearing in Eq. 4.7 is now replaced by a Lorentzian:

$$\langle x_i | e^{-\Delta t \Gamma |p|} | x_{i+1} \rangle = \frac{1}{\pi} \frac{\Delta t \Gamma}{(\Delta t \Gamma)^2 + (x_i - x_{i+1})^2} \,. \tag{4.11}$$

With a certain effort, we have succeeded in generalizing the bisection algorithm to implement in a very effective way the dynamics provided by this Lorentzian propagator, see App. C.12.

We present here the results of a PIMC-QA approach which implements the relativistic kinetic energy, through a bisection algorithm with such a "Lorentzian move". In Fig. 4.3 we report the residual energy results for the case of the V_{asym} potential (Top) and V_{sym} (Bottom). The results are obtained using $P = 40$ Trotter's slices and $l = 2$ bisection steps. Averages and statistical errors are calculated, as usual, with 10^3 repetitions of every annealing run. It is quite clear that – as in the classical case – the Lorentzian move (*alias*, in the present context, the relativistic kinetic energy) greatly speeds up the QA behavior for the difficult V_{asym} case. As usual, the V_{sym} case does not show any big qualitative change.

4.5 Summary and discussion

As a way of summary, we would like to briefly stress some of the major points touched upon in this chapter.

1. *Tunneling and sampling difficulties.* A tunneling event, and its associated LZ crossing, can cause severe difficulties to an allegedly "state-of-the-art" PIMC-QA algorithm. The difficulty is associated to a poor sampling of the action which does not accurately describes the rare, but all-important, tunneling events. This problem has been cured by the ad-hoc introduction of "instanton moves", but this cure, while instructive, is playing an unfair game, i.e., using detailed information on the potential landscape we want to optimize. The generalization of these "instanton moves" to more complicated potentials, let alone to generic hard optimization problem, is basically not available: it would require, among other things, a precise notion of the location of the minima were tunneling occurs.

2. *Other limitations of PIMC.* PIMC-QA suffers from a certain number of other limitations. First of all, it works with a *finite temperature* T, and this sets up a thermal energy lower bound below which we cannot possibly go (this limit was particularly clear in our double-well example). Second, a large number of Trotter slices P can cause additional sampling problems in that an effective decorrelation of the configurations becomes harder and harder, even if a multi-step bisection algorithm is employed. Moreover, the *Trotter break-up* itself, see Eq. 4.4, can cause difficulties whenever, for generic kinetic energy T, the form of the free propagator, generalizing the simple Gaussian result of Eq. 4.7, is now known analytically. This was indeed the difficulty met in the Traveling Salesman Problem (TSP) 1.

3. *Role of kinetic energy.* The choice of the kinetic energy is clearly all important in QA: Sec. 4.4, illustrating the improvements in PIMC-QA upon using a relativistic kinetic energy, is particularly instructive. Notice that, in order to sample a probability distribution required by the (one-dimensional) Lorentzian-bisection scheme (see App. C.12), we have devised a rejection sampling technique which has 50% acceptance rate. [8] This is the only feasible implementation presently known. A trivial generalization of our algorithm in D dimensions requires a CPU time at least D times bigger than for the one-dimensional case. This is definitely a good theoretical scaling, although the convenience of such a scheme is not *a priori* obvious and further tests on higher dimensional problems would be needed.

In view of the previous points, it is natural to explore alternative Quantum Monte Carlo schemes for performing stochastic implementations of QA. One such scheme, which in principle does not suffer from many of the limitations of PIMC, is the Green's Function Monte Carlo scheme, which we discuss in the next chapter, with an application to Ising spin glasses.

[8] This is at variance with the Gaussian case, where one can sample the distribution exactly, without rejection.

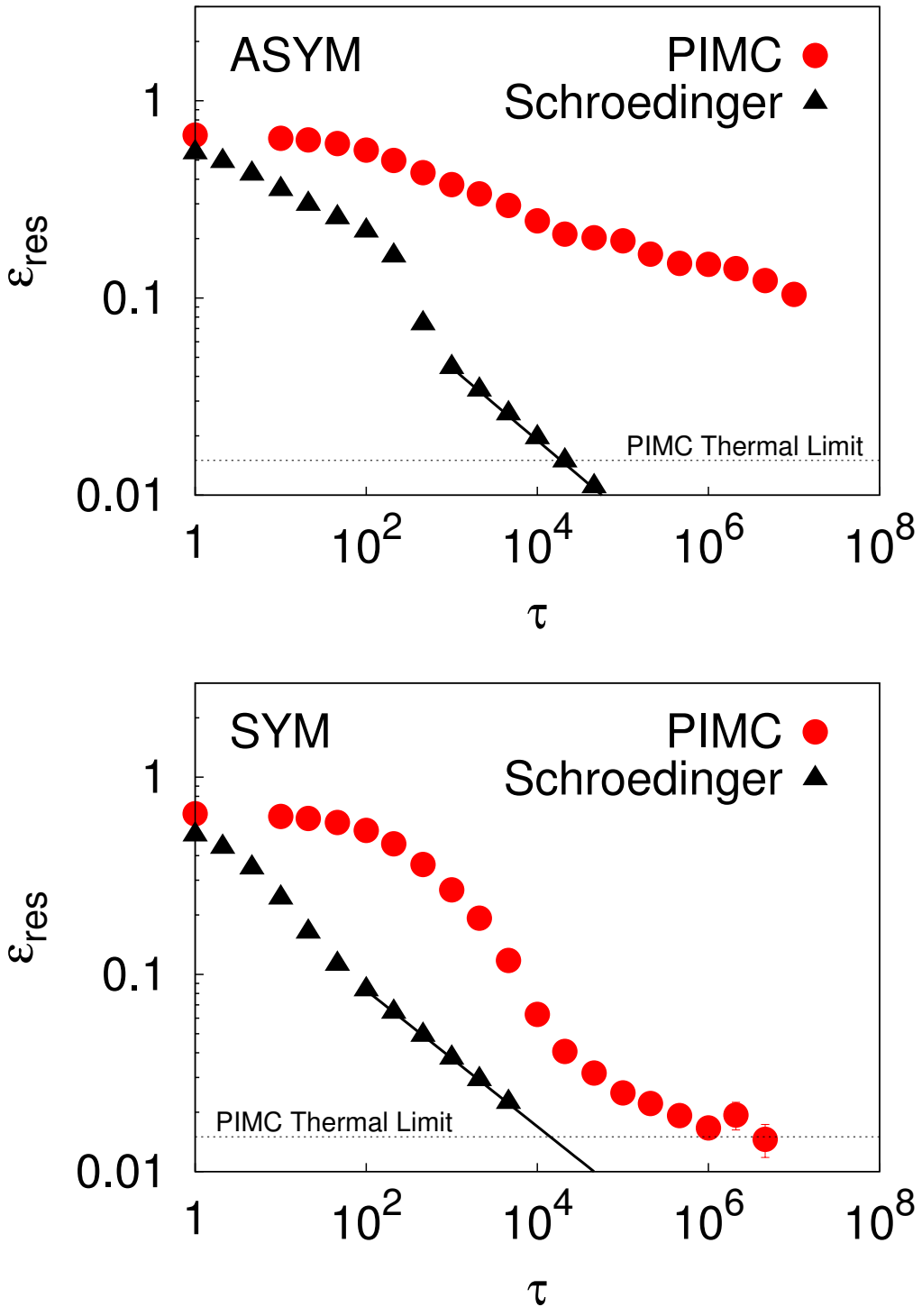

Figure 4.1: (Top) PIMC-QA residual energy for the V_{asym} potential, using a fourth-order action and the bisection algorithm. The dashed line indicates the thermal equipartition limit $T/2$, for $T = 0.03$. As a reference, the results obtained by exact integration of the imaginary-time Schrödinger equation (see Chap. 2) are reported. (Bottom) Same as above, for a symmetric potential V_{sym}.

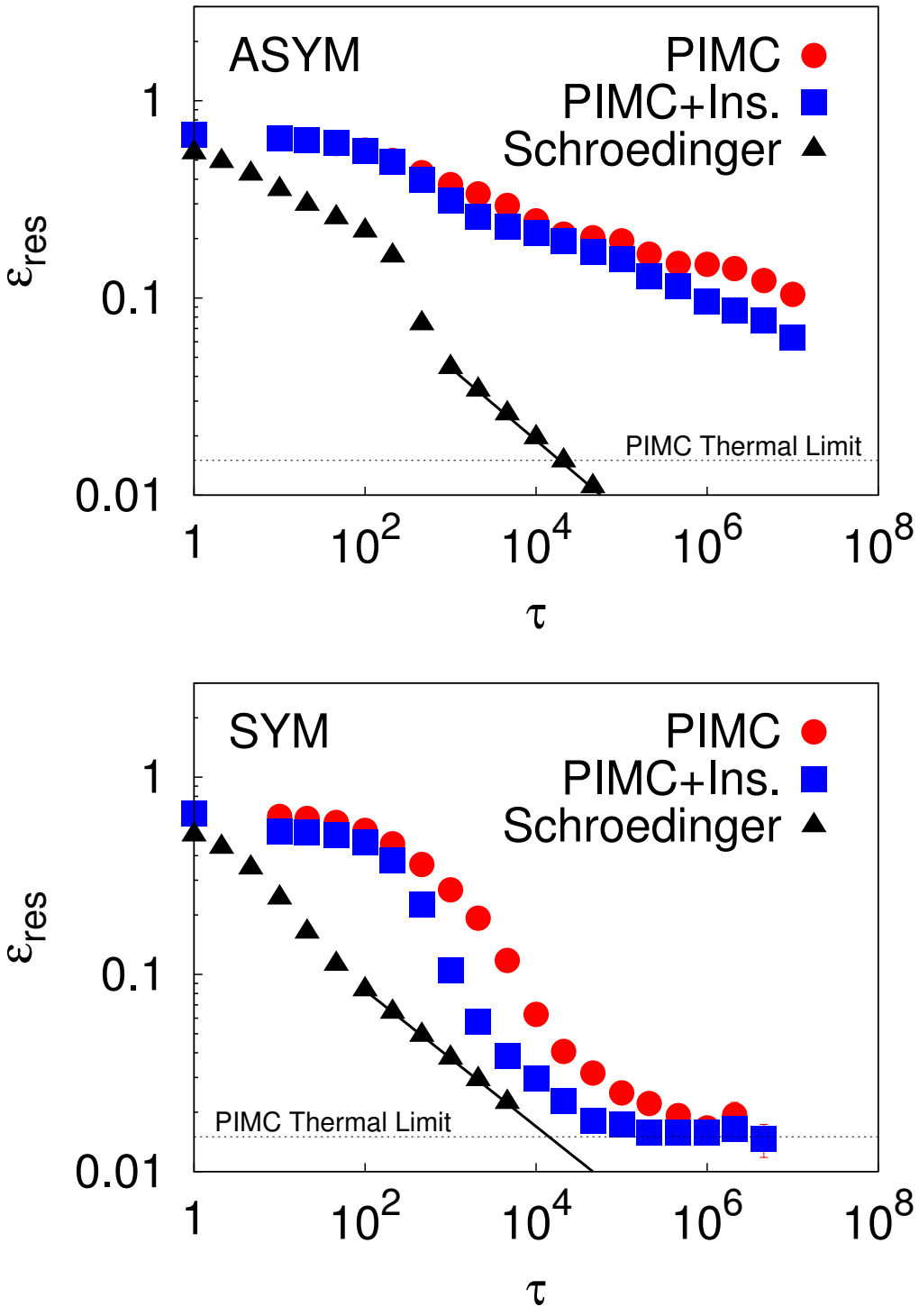

Figure 4.2: (Top and Bottom) Same as in Fig. 4.1, with the instanton move allowed. As a reference the results obtained without instanton move are still reported.

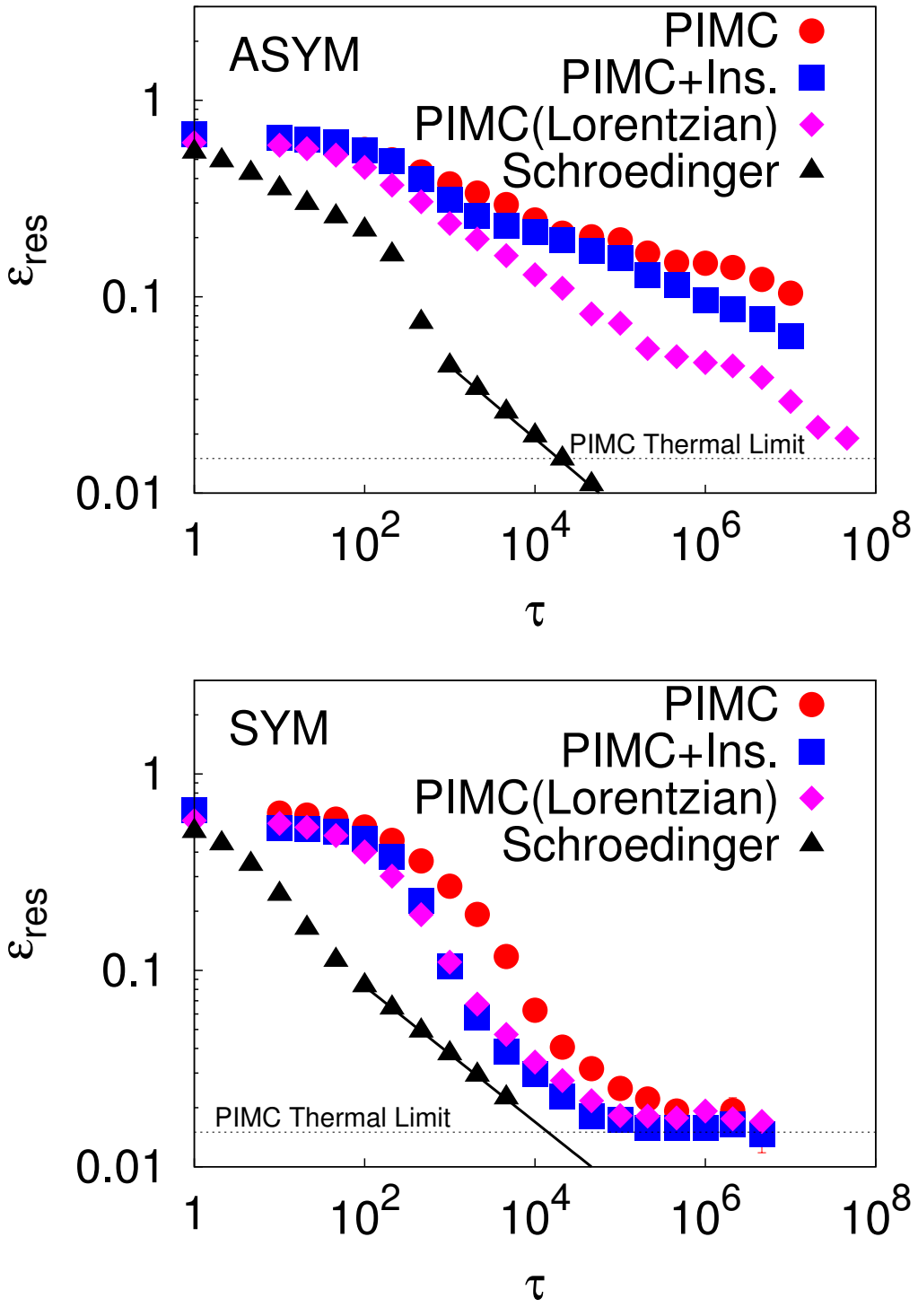

Figure 4.3: Same as in Fig. 4.1, but for a QA based on the relativistic kinetic energy of Eq. 4.10, implemented via a bisection algorithm adapted to Lorentzian moves. As a reference, the results obtained by exact integration of the imaginary-time Schrödinger and by the previous Gaussian-based PIMC-QA with and without instanton move are still reported.

Chapter 5

Annealing by Green's Function Monte Carlo

In chapter 4 we studied a Path-Integral Monte Carlo (PIMC) based quantum annealing (QA) algorithm, focusing our attention on a simple, but highly instructive, toy-problem: the double well potential. There we learned a few potential dangers of PIMC-QA, in particular: i) the unavoidable finite temperature T used in the simulation, which provides a thermal lower limit to the average residual energies attained by the algorithm; ii) the possible sampling difficulties (ergodicity breaking) of the PIMC action associated to Landau-Zener tunneling phenomena.

We propose here to investigate an alternative Quantum Monte Carlo (QMC) based QA algorithm. A very natural choice for this is provided by the *Green's Function Monte Carlo* (GFMC). [1] As for the system on which to test this new GFMC-QA algorithm, a very natural benchmark problem is provided by the *Ising spin-glass* ground state search, a challenging optimization problem already addressed through QA in the recent past [3], particularly through PIMC-QA [10, 11]. We refer the reader to the Chap. 1, for an illustration of the previous PIMC-QA results on the problem. Here we simply recall that the Hamiltonian of the problem reads:

$$H = -\sum_{\langle i,j \rangle} J_{i,j}\, \sigma_i^z \sigma_j^z - \Gamma \sum_i \sigma_i^x = H_{cl} + H_{kin} \ , \tag{5.1}$$

where $\sum_{\langle i,j \rangle}$ indicates a sum over nearest-neighbors, $J_{i,j}$ are random nearest-neighbor

[1] Depending on details of the method, Diffusion Monte Carlo (DMC) is another name for a similar approach.

Ising coupling constants, and σ_i^z, σ_i^x are Pauli's matrices on site i. More precisely, the problem instance we concentrate on is the one which has been analyzed in detail in Ref. [10, 11], which refers to a two-dimensional (2D) case, with an $L \times L$ lattice (with $L = 80$), and the J_{ij}'s drawn from a flat distribution in $(-2, 2)$. [2]

If we denote by $\{S_i\}$ a generic configuration in the Hilbert space (where $S_i = \pm 1$ are the eigenvalues of σ_i^z), the classical function we want to minimize is just given by the first term in Eq. 5.1, i.e., $E_{cl}(\{S_i\}) = \langle \{S_i\} | H_{cl} | \{S_i\} \rangle$ is our *potential energy*. The non-trivial nature of the minimization task is due to the fact that the number of configurations possible for a finite-size lattice with $N = L^2$ sites is 2^N, an astronomical number when $N = 80^2 = 6400$, while frustration and disorder in the couplings J_{ij} leave no simple rule for constructing good energy configurations. The second term in Eq. 5.1, $-\Gamma \sum_i \sigma_i^x$, is the needed source of quantum fluctuation, which plays therefore the role of *kinetic energy*. The transverse field Γ represents therefore the annealing parameter of the system; as usual, the goal is to follow the time-dependent dynamics given by the Hamiltonian in Eq. 5.1 with a $\Gamma(t)$ which starts from very large values, and vanishes in a certain annealing time τ.

We emphasize that such a transverse field term is not just a theory dream. Indeed, the whole field of QA was strongly revived by experimental results on the disordered Ising ferromagnet $LiHo_{0.44}Y_{0.56}F_4$, where the transverse field Γ was actually applied to the system and manipulated in the laboratory, to perform the first true *quantum annealing experiment* [4, 62].

The rest of the chapter is organized as follows. We begin by describing (Sec. 5.1) the main ideas behind the GFMC approach and the results of a variational study of two possible trial wavefunctions (Sec. 5.2). We then move to discussing the results obtained by GFMC for fixed values of Γ (Sec. 5.3). Finally, we present (Sec. 5.4) the GFMC-QA results.

[2] The choice of a 2D case is motivated by the fact that the Ising glass ground state search is actually a *polynomial problem* in 2D, and very efficient *branch and cut* algorithms (see Ref. [35] and the URL `www.informatik.uni-koeln.de/ls_juenger/projects/sgs.html`) are know to find the true optimal state energy E_{GS}, which thus provides a clear benchmark for any annealing study. Yet, the problem has a prohibitively large complexity for any direct application of a deterministic physical dynamics (classical or quantum), like a true Glass [44, 61].

5.1 Introduction to GFMC

We give here an account of the main ideas behind a GFMC algorithm. For a more detailed introduction to the method, see App. D.1 and Ref. [63].

The Green's Function Monte Carlo (GFMC) is a stochastic technique which is based on a rather simple but important property of powers of matrices, strongly reminiscent of the discussion of Markov chains and the associated Master equation given in Chap. 3. We start by explaining the key idea in a deterministic context. Suppose we want to get the exact ground state wavefunction $\phi_0(x)$ of a certain Hamiltonian H, with corresponding eigenvalue E_0. We start from an initial state $\psi_0(x)$ with non-zero overlap with the exact ground state ϕ_0 (e.g., a randomly generated vector in the Hilbert space, which is typically never orthogonal to the exact ground state). The exact ground state can be *filtered out* from ψ_0 by iteratively applying to the state powers of $1-\Delta t H$, where $\Delta\, t$ is chosen such that (in absolute value) $|1-\Delta t E_0|$ is the *largest eigenvalue* of $|1-\Delta t H|$. [3] More precisely, we define, recursively:

$$\psi_{n+1}(x') = \sum_x \langle x'|1-\Delta t H|x\rangle \psi_n(x) = \sum_x (\delta_{x',x} - \Delta t H_{x',x})\psi_n(x)\ , \tag{5.2}$$

with ψ_0 the initial state selected. It is simple to show that, for large n, the iterated state ψ_n converges to the ground state $\phi_0(x)$, if Δt is small enough that the condition $\max_i|1-\Delta t E_i| = |1-\Delta t E_0|$ is satisfied, and there is a finite gap between E_0 and E_1. Indeed, we can expand the initial wavefunction in the basis of eigenstates ϕ_i of H of energies E_i, $|\psi_0\rangle = \sum_i a_i|\phi_i\rangle$ with $a_i = \langle\phi_i|\psi_0\rangle$, and $\sum_i a_i^2 = 1$. Thus, at iteration n, we have:

$$|\psi_n\rangle = \sum_i a_i(1-\Delta t E_i)^n|\phi_i\rangle\ . \tag{5.3}$$

It is then clear that, by increasing n, the ground state component in the expansion (5.3) will grow much faster than all the excited state components $|\phi_{i>0}\rangle$, provided $|1-\Delta E_0| > |1-\Delta E_i|$ for $i > 0$. Finally, apart for an irrelevant normalization factor $(1-\Delta E_0)^n$, the convergence of ψ_n to ϕ_0 is obtained with an expo-

[3] Notice that many operator functions of the Hamiltonian would do the job as well. The most important one is obviously the imaginary-time propagator $e^{-\Delta t H}$, which is however rather difficult to implement, due to the exponential. Another possible choice is given by $\Lambda - H$ where Λ is a suitable shift constant which takes care of making $|\Lambda - E_0| > |\Lambda - E_i|$. What we actually use is $1-\Delta\, t\ (H-\epsilon_T)$, where ϵ_T is an estimate of E_0.

nentially increasing accuracy in n, namely with an error proportional to Δ^n, where $\Delta = \max_{i\neq 0}|1-\Delta t E_i|/|1-\Delta t E_0|$.

The method just described is called *iterated power method*, and is used in many context in condensed matter and statistical mechanics. In the description given above, all we have to do is to repeatedly apply a matrix, $1-\Delta t H$, to a vector, ψ_n, an operation which is very straightforward if the Hilbert space of the problem considered is sufficiently small. "Small" here means that a few million elements can still be managed, but not much more: on the contrary, a "small" Ising problem on a 32×32 lattice has a Hilbert space with $2^{1024}\approx 10^{308}$ elements, and the application of the power method as a matrix-vector product is totally out of question.

Let us see, now, how a *stochastic implementation* of this idea works. We denote, as usual, by x the discrete labels specifying a given state of the Hilbert space of our system (for instance, specifying all the electron positions and spins, for an electronic problem, or just a spin configuration, for a spin problem). We will also assume that, given the Hamiltonian H, the matrix elements $\langle x'|H|x\rangle = H_{x',x}$, for given x, can be computed efficiently for each x'. Typically, for a lattice Hamiltonian, though the dimension of the Hilbert space increases exponentially with the system size L, the number of non-vanishing entries of the matrix representing H, $H_{x',x}\neq 0$, is quite small – of the order of L –. Consider the type of problem required by the iterated power method:

$$\begin{aligned}\psi_{n+1}(x') &= \sum_x G_{x',x}\psi_n(x)\\ G_{x',x} &= \langle x'|1-\Delta t H|x\rangle = \delta_{x',x}-\Delta t H_{x',x}\ ,\end{aligned} \tag{5.4}$$

where we have defined $G_{x',x}$, the so-called *Green's function*, as the matrix to be applied to the state. The problem in Eq. 5.4 looks superficially similar to the Master equation introduced in Chap. 3 (see, in particular, Eq. 3.1) with a few very *crucial differences*: i) the $\psi_n(x)$ are *not probabilities*, but *amplitudes*, and as a consequence we do not certainly have $\sum_x \psi_n(x)=1$; ii) the Green's function $G_{x',x}$, unlike the transition probability of the Master equation, is not necessarily made of non-negative elements, and is, in general, not column-normalized. In summary, the process underlying the iterated-power method is *not* a properly defined Markov chain, and, therefore, we cannot immediately use a Monte Carlo approach.

Problem ii) above can be quite serious: if some of the matrix elements of $G_{x',x}$ are *negative*, no possible interpretation of it as a "transition probability" is possible. This

is at the hearth of the so-called *sign problem*, which plagues many Quantum Monte Carlo approaches. In the following, therefore, we will assume that a choice of the basis is possible in which *no sign problem* exist, i.e., all matrix elements of G are non-negative, $G_{x',x} \geq 0$. [4] Still, we miss the correct column-normalization,

$$\sum_{x'} G_{x',x} = 1 - \Delta t \sum_{x'} H_{x',x} \stackrel{def}{=} b_x \neq 1 \ , \tag{5.5}$$

and we have to circumvent this difficulty. A way out, is to factorize G in terms of a stochastic matrix $p_{x',x}$ – by definition, a matrix with all positive elements $p_{x',x} \geq 0$, and with the normalization condition $\sum_{x'} p_{x',x} = 1$ for all columns – times the scale factor b_x defined above. Indeed, with the previous definition (5.5) of b_x, the matrix $p_{x',x} = G_{x',x}/b_x$ is trivially positive and column normalized, and is therefore the stochastic matrix we are looking for. In summary, we have split G into:

$$\begin{aligned} G_{x',x} &= p_{x',x} b_x \\ p_{x',x} &= G_{x',x}/b_x = G_{x',x} / \sum_{x'} G_{x',x} \ . \end{aligned} \tag{5.6}$$

$p_{x',x}$ would be a suitable transition matrix for a Markov chain in x-space, but evidently this is not enough for our purposes, because all the information about b_x, which is in some way related to the wavefunction amplitude, is not contained in it. The crucial idea is now to *extend* the configuration space in which the Markov process is defined, adding to the x a non-negative weight factor w: this extended configuration space is then labeled by (x, w), where x runs as usual on the Hilbert space configurations, and $w \geq 0$. The pair (x, w) is often called a *walker*, because it will be our basic entity in the Markov chain "random walk". The weight part will take care of b_x, while x will be taken care of by $p_{x',x}$. More precisely, if (x_n, w_n) indicates a walker at iteration time n, in this extended configuration space, we set up the following Markov process: [5]

a) generate $x_{n+1} = x'$ with probability p_{x',x_n}

[4] This is certainly true for the Ising glass in a transverse field. More generally, since the choice of the kinetic energy to be used in QA is at our disposal, it is wise to choose the signs in H_{kin} (see Eq. 5.1) such that no sign problem occurs.

[5] Practically, given x_n at iteration n, in order to define x_{n+1}, it is enough to subdivide the interval $(0, 1)$ into smaller sub-intervals of lengths p_{x',x_n} for all possible x' connected to x_n with non-zero probability p_{x',x_n}, which are of order L in number. Then, a pseudo-random number ξ between 0 and 1 is generated. This ξ will fall inside one of the above defined sub-intervals, with a probability of hitting

$$\text{b) update the weight with } w_{n+1} = w_n b_x \ . \tag{5.7}$$

More formally, the Markov process is defined by the following conditional probability $K(x_{n+1}, w_{n+1}|x_n, w_n)$, i.e., the transition probability for obtaining a new walker (x_{n+1}, w_{n+1}), at iteration $n+1$ given the old one (x_n, w_n), at iteration n:

$$K(x_{n+1}, w_{n+1}|x_n, w_n) = p_{x_{n+1},x_n}\delta(w_{n+1} - w_n b_{x_n}) \ . \tag{5.8}$$

Thus, the Master equation corresponding to the probability density $P_n(x_n, w_n)$ in the enlarged configuration space is given by (see Eq. 3.1): [6]

$$P_{n+1}(x', w') = \sum_x \int dw \, K(x', w'|x, w) \, P_n(x, w) \ . \tag{5.9}$$

In words: the walker performs a *random walk* in the Hilbert space x of the system and in the weight space w; such a random walk is composed of a standard Markov chain in x-space, associated to the $p_{x',x}$, plus a *multiplicative process* for the weight $w_n \to w_{n+1} = w_n b_x$. By moving in this way, the walkers visits every point in the (x, w)-space with a probability $P_n(x_n, w_n)$. The crucial question is: how is $\psi_n(x)$ related to $P_n(x_n, w_n)$? It is relatively simple to show that, if we define the wavefunction amplitude $\psi_n(x)$ to be the first moment of $P_n(x, w)$ with respect to the weight w,

$$\psi_n(x) =< w_n \delta_{x,x_n} > \stackrel{def}{=} \sum_{x_n} \int dw_n \, w_n \, \delta_{x,x_n} P_n(x_n, w_n) = \int dw_n \, w_n \, P_n(x, w_n) \ , \tag{5.10}$$

then the basic iterative equation (5.4) is perfectly satisfied. [7] Needless to say, one can construct a formal proof of the convergence of the Markov process in much the same way as is usually done for the standard Markov chain [63].

---

the sub-interval corresponding to a certain x' which exactly equal to p_{x',x_n}. This clearly defines x_{n+1} according to the desired Markov chain (5.7). The walker weight w_n is then automatically updated in a multiplicative way, as $w_n \to w_{n+1} = w_n b_{x_n}$, and the new walker configuration (x_{n+1}, w_{n+1}) is completely defined.

[6] Here and in the following, the integration limits over the variable w are assumed to run over the whole range $-\infty, \infty$, the probability density $P(x, w)$ being *zero* for the values of w that are not allowed by the Markov chain (e.g., $w < 0$).

[7] The average $< \cdots >$ appearing in Eq. 5.10 is intended over infinitely many independent realizations of the Markov chain, distributed according to $P_n(x_n, w_n)$, and is rarely pursued in actual practice, since there is too much information contained in $\psi_n(x)$ for every configuration x in the Hilbert space. The actual goal is to calculated a few physically important averages, like energy and correlation functions, and not the actual $\psi_n(x)$ itself.

Summarizing, the power method guarantees convergence, for large n, to the ground state of H and thus the walker (x_n, w_n) gives information on the ground state wavefunction $\psi_{GS}(x)$, namely

$$\psi_n(x) =< w_n \delta_{x,x_n} > \stackrel{n\to\infty}{\longrightarrow} \psi_{GS}(x) \ .$$

On the other hand, the well defined Markov process behind the whole construction allows, *in principle*, to follow the iterations *stochastically*, i.e, through a Monte Carlo procedure.

Local Energy and Ground State Energy. The ground state energy of the problem, can in principle be extracted from an appropriate average of the so-called *local energy*:

$$e_L(x) = \sum_{x'} H_{x',x} \ . \tag{5.11}$$

More precisely, if we take the ratio between the averages of the two random variables $w_n e(x_n)$ and w_n over the Markov chain (5.7), and we use Eq. 5.10 and Eq. 5.11, we easily obtain:

$$\frac{< w_n e_L(x_n) >}{< w_n >} = \frac{\sum_{x_n} \int dw_n w_n e_L(x_n) P_n(x_n, w_n)}{\sum_{x_n} \int dw_n w_n P_n(x_n, w_n)} = \frac{\sum_{x_n} e_L(x_n)\psi_n(x_n)}{\sum_{x_n} \psi_n(x_n)} \stackrel{n\to\infty}{\longrightarrow} E_{GS} \ . \tag{5.12}$$

Indeed, for large n, when $\psi_n(x)$ approaches the ground-state wavefunction $\psi_{GS}(x)$ (up to a normalization constant) the numerator will converge exactly to the ground state energy E_{GS} times the denominator.

Importance Sampling. According to the GFMC scheme presented until now, the fluctuations in the energy calculation, Eq. (5.11) and Eq. 5.12, are independent of any possible knowledge we have of the ground state. Quite often, however, one has a reasonably good guess of the ground state wavefunction $\psi_{GS}(x)$, possibly dependent on external parameters that have been optimized in some way through a Variational Calculation. It would be desirable to use such information in reducing the fluctuations in the averages we calculate. This is indeed possible through a slight modification of the GFMC proposed until now, which goes under the name of *importance sampling*. Suppose we have a reasonable guess of $\psi_{GS}(x)$ in the form of some nodeless [8] wavefunction $\psi_T(x)$, known as *trial (or guiding) wavefunction*. It is then enough to

[8] Due to the fact that the Green's function is non-negative, and hence the Hamiltonian has non-positive off-diagonal elements, the Perron-Frobenius theorem [49][page 53ff] guarantees that the Ground State wavefunction must be unique and positive.

substitute G with the so-called *importance sampling* Green's function:

$$\bar{G}_{x',x} = \psi_T(x') G_{x',x} \psi_T^{-1}(x) \ . \tag{5.13}$$

Notice that, in general, $\bar{G}_{x',x}$ is *no longer symmetric*. Whenever $\bar{G}_{x',x} \geq 0$ for all (x', x), it is possible to apply the same decomposition in (5.6) to $\bar{G}$, defining the corresponding Markov chain (5.7) with:

$$\begin{aligned} p_{x',x} &= \bar{G}_{x',x}/b_x \\ b_x &= \sum_{x'} \bar{G}_{x',x} = 1 - \Delta\, t\, \frac{\sum_{x'} \psi_T(x') H_{x',x}}{\psi_T(x)} = 1 - \Delta\, t\, \bar{e}_L(x) \ , \end{aligned} \tag{5.14}$$

where the last equality defines the importance-sampling local energy $\bar{e}_L(x)$. Quite amusingly, if the guessed trial wavefunction ψ_T coincides with the actual ground state wavefunction $\psi_T(x) = \psi_{GS}(x)$, then $\bar{e}_L(x) = E_{GS}$ is a constant, and statistical fluctuations in the calculations of Eq. 5.12 *vanish exactly*. Therefore, by variationally improving the quality of the guiding function $\psi_T(x)$ we can substantially reduce the error bars in the energy calculation.

Fluctuations go wild: Many walkers and Branching. The scheme we have just proposed works only *in principle*, but, unfortunately, *not in practice*. The reason for this failure is quite simple to understand. While $x_n \to x_{n+1}$ is a honest Markov process, the weight update $w_n \to w_{n+1} = w_n b_{x_n}$ is a *multiplicative* process with random factors b_{x_n}, which is prone to very large fluctuations: w_n might grow tremendously large, or small, in just a few iterations of the process, and the whole procedure would go wild, because error bars in the calculations of the averages grow in an uncontrolled way. The cure to this disease goes through the introduction of *many walkers* and through performing occasional "reconfigurations" of their weights, via the so-called *branching*. In practice, one propagates simultaneously a set of M walkers defined by weights w_i and configurations x_i, for $i = 1, \cdots M$. Before the variance of the weights w_i becomes too large, one appropriately redefines the set of walkers in such a way as to drop out those with a too small weight, and to generate copies of the more important ones. [9] After the reconfiguration (branching) all the walkers are

[9] The branching is just a particular Markov process applied to the configurations (x_j, w_j), which leads to new walkers (x'_j, w'_j). Each new walker (x'_j, w'_j), with $j = 1 \cdots M$, will have the same weight $w'_j = \bar{w} = \sum_j w_j/M$ and an arbitrary configuration x'_j picked up among the M possible old ones x_k, $k = 1 \cdots M$, with a probability p_k proportional to the weight of that configuration, $p_k = w_k / \sum_j w_j$.

given again the same weight, and can be independently propagated again for a certain number of iterations. By iterating this process, the weights of all the walkers will be kept approximately equal during the simulation. This property yields a considerable reduction of statistical errors, as the variance of the average weight $\bar{w} = \sum_i w_i/M$ is reduced by a factor $\propto \sqrt{M}$. One can show [63] that such a reconfiguration process does not modify, on average, the relevant wavefunction $\psi_n(x)$.

Continuous-time formulation.[64, 65] In the limit $\Delta t \to 0$, clearly, the diagonal elements dominate the Green's function $\bar{G}$, since off-diagonal elements $\bar{G}_{x',x}$ $(x' \neq x)$ have a factor Δt in front: if we were to apply the algorithm in a straightforward way, in most of the iterations we would generate a diagonal move, $x_{n+1} = x_n$, which is highly inefficient. We can calculate, however, what is the total escape rate (i.e., probability per unit time) from a given configuration x:

$$\lambda_x = \lim_{\Delta t \to 0} \frac{1}{\Delta t} p_{esc} = \sum_{x' \neq x} \psi_T(x') \langle x'|(-H)|x\rangle \psi_T^{-1}(x) \ .$$

In the limit $\Delta t \to 0$, the distribution of escape times from a given configuration x is Poissonian [66][page 31ff]:

$$P_x(t) = \lambda_x e^{-\lambda_x t} \ . \tag{5.15}$$

To see this explicitly, notice that the probability of escaping from x exactly during iteration n (i.e., in the time interval $[t - \Delta t, t]$, where $t = n\Delta t$) is simply $(1 - p_{esc})^{n-1} p_{esc}$. Therefore:

$$\Delta t P_x(t) = (1 - \Delta t \lambda_x)^{n-1} \ (\Delta t \lambda_x) \ ,$$

so that taking the limit $n \to \infty$, $\Delta t \to 0$, with $t = n\Delta t$ *fixed*, we easily recover Eq. (5.15). During the $n-1$ diagonal moves prior to the n-th off-diagonal one, the weight of a walker has to be repeatedly multiplied ($n-1$ times) by $b_x = 1 - \Delta t \bar{\epsilon}_L(x)$. In the continuous-time limit this implies that the weight is multiplied as follows:

$$w \to w' = w e^{-t\bar{\epsilon}_L(x)} \ . \tag{5.16}$$

Therefore, it is enough to sample the average time t we stay in a given configuration x according to $P_x(t)$ (which we do very easily [66][page 31ff] by generating a random number ξ in $(0, 1]$ and taking $t = -\ln(\xi)/\lambda_x$), and directly go on selecting an off-diagonal move according to $\bar{G}_{x',x}$. In other words, there is no need to *try* to move out

of a configuration x: if we can estimate the average time we need to stay in x, we can simply *advance the clock* by t, and exit from x right away. Each Monte Carlo step, therefore, is associated to an off-diagonal move.

5.2 The choice of the trial wavefunction.

An appropriate *importance sampling*, via a good trial wavefunction $\psi_T(x)$, is in most cases a crucial ingredient for a successful GFMC approach. Unfortunately, finding a good trial $\psi_T(x)$ for the Ising spin-glass in a transverse field is far form trivial. Frustration and disorder in the couplings J_{ij} induce, evidently, non-trivial correlations between the spins, which are quite hard to catch. We illustrate below two simple choices for ψ_T which we have explored in the present work.

Single-site product wavefunction. As a first choice for a trial wavefunction, we disregard all possible inter-site correlation and consider, in a *mean field* spirit, a ψ_T which is a simple product of single-site wavefunctions:

$$|\psi_T^{(MF)}\rangle = \prod_{i=1}^{N} \left(\frac{e^{+\frac{h_i}{2}} \mid \uparrow\rangle_i + e^{-\frac{h_i}{2}} \mid \downarrow\rangle_i}{\sqrt{2\,\cosh(h_i)}} \right) , \tag{5.17}$$

where $\{h_i\}$ are local fields on each site i, which we use as variational parameters to be optimized. In terms of the *local magnetization* $m_i = \tanh(h_i)$, the average energy reads:

$$E_T^{(MF)}(\{h_i\}) = \langle \psi_T^{(MF)} | H | \psi_T^{(MF)} \rangle = -\sum_{\langle i,j \rangle} J_{i,j}\, m_i\, m_j - \Gamma \sum_i \sqrt{1 - m_i^2} , \tag{5.18}$$

and the stationary conditions required by the minimization read:

$$\frac{\partial\, E_T^{(MF)}}{\partial\, h_i} = -\,(1 - m_i) \left(\sum_{j \in N(i)} J_{i,j}\, m_j \right) + \Gamma\, m_i \sqrt{1 - m_i^2} = 0 \quad \forall i , \tag{5.19}$$

where $N(i)$ indicates the nearest-neighbors of site i. When Γ is large, the obvious solution to the Eq. 5.19 is obtained by taking $m_i = h_i = 0$ for all sites i: this is nothing but the $\Gamma = \infty$ exact solution, with all spins aligned along $+\hat{x}$. Such a large-Γ quantum paramagnetic phase, with a gap of order 2Γ separating the ground state from higher excited states, survives down to some critical value Γ_{cr}, after which non-trivial solutions of Eq. 5.19 with non-vanishing local magnetizations $m_i \neq 0$ start to

appear. [10] In this small-Γ region, in which solutions with $m_i \neq 0$ start to appear, our minimization problem is just the quantum counterpart of the well known Weiss mean-field approach for the classical random Ising model [68], which is known to get into big difficulties in the $T < T_g$ *glassy phase*. [11] We anticipate that similar difficulties will plague our search for the minima of Eq. 5.18 in the small-Γ phase. Essentially, the minimization problem is characterized by the presence of many local minima: searching for the minima of Eq. 5.18 at small Γ is not much simpler, in essence, than solving the original classical ground state search problem, except that the configuration space in which we search for the minima is continuum, $\{m_i \in \mathbf{R}\}$, instead of discrete, $\{S_i = \pm 1\}$.

Boltzmann-like wavefunction. Another quite natural choice of trial wavefunction is a Boltzmann-like function of the form:

$$\langle \{S_i\} | \psi_T^{(\beta)} \rangle = \mathcal{N}(\beta)\, e^{-\beta^2 E_{cl}(\{S_i\})} \ , \tag{5.20}$$

where β^2 plays the role of an inverse effective temperature, $E_{cl}(\{S_i\})$ is the classical energy of a given configuration $\{S_i\}$, and $\mathcal{N}(\beta)$ is a normalization factor. Here β is our variational parameter. Once again, for large Γ we expect to find $\beta = 0$ (the exact $\Gamma = \infty$ solution), while, by decreasing Γ, larger values of β will privilege regions where the "potential energy" $E_{cl}(\{S_i\})$ has local minima.

Variational results. We start by illustrating the results obtained for the first choice of trial wavefunction, see Eq. 5.17. Given the explicit analytic expression of the energy to be minimized, Eq. 5.18, we have used a Conjugate Gradients (CG) algorithm to find the optimal values for the $\{h_i\}$ and the corresponding optimal value of the variational energy per spin, $\epsilon_{var}^{(MF)} = \min_{\{h_i\}} E_T^{(MF)}(\{h_i\})/N$. However, as anticipated, the results obtained for $\epsilon_{var}^{(MF)}$ when $\Gamma < \Gamma_{cr}$ depend on the initial point provided to the algorithm, so that the only meaningful thing to do is to show *histograms* of $\epsilon_{var}^{(MF)}$ obtained by repeated CG-minimization searches. In Fig. 5.1 we report the results of four calculations for $\Gamma = 0.1$, which turns out to be in the low-Γ glassy phase. The two right panels show the histogram of 50 repeated searches for the CG minimum,

[10] The value of Γ_{cr} is not easy to establish even within our variational framework. For the case of a square lattice ordered ferromagnet ($J_{i,j} = J$), it is instead simple to work out that $\Gamma_{cr} = 4\,J$. For QMC results on the disordered case, see for instance Ref. [67].

[11] In the classical case of finite T, even improvements of the Weiss approach, through the so-called TAP equations – where the cavity field created at site i by all other spins is more properly calculated – are known to fails, essentially, for $T < T_g$.

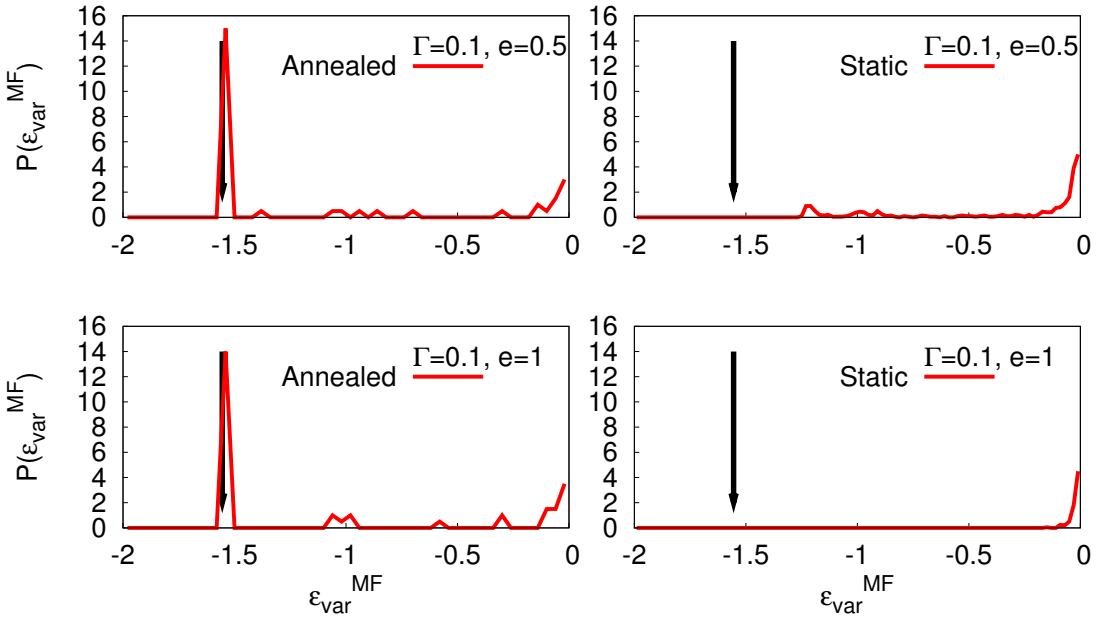

Figure 5.1: Distribution of the variational energy per spin $\epsilon_{var}^{(MF)} = Min_{\{h_i\}}\langle\psi_T^{(MF)}|H|\psi_T^{(MF)}\rangle/N$ obtained by optimizing, with a Conjugate Gradients technique, the local-fields h_i in the trial wavefunction $|\psi_T^{(MF)}\rangle$ defined in Eq. 5.17. These results refer to the case of a fixed transverse field $\Gamma = 0.1$. The data in the two left panels have been obtained by using a variational annealing procedure (see text), while the data in the right panels have been obtained by a straight conjugate gradient minimization. In both cases, the starting point is a randomly distributed choice $h_i \in (-e, +e)$, with $e = 0.5$ in the upper panels and $e = 1$ in the lower ones. The histograms are obtained by dividing the energy interval $[-2, 0]$ into 100 subintervals.

starting from initial h_i which are randomly distributed in the interval $(-e, +e)$, with $e = 0.5$ (top panel) or $e = 1$ (bottom panel). The big solid arrow marks the location of the classical ground state energy per spin, $\epsilon_{GS} = E_{GS}/N \approx -1.5805167$, as a reference. Although, in principle, quantum effects due to the finite value of Γ make $\epsilon_{var}^{(MF)} \neq \epsilon_{GS}$, the small Γ value used ($\Gamma = 0.1$) does not justifies such a large difference between the two quantities: in most of the attempts we simply find minima with $\epsilon_{var}^{(MF)} \sim 0$, that have nothing to do with the classical GS which quite likely dominates the physics for such a low Γ value. Only rarely, see top right panel, we end up with minima in the range $\epsilon_{var}^{(MF)} \sim -1.25$, still quite far from ϵ_{GS}. In the two left panels of

Fig. 5.1 we show the results of a sort of "variational" QA CG-based minimization. [12] In essence, we started from the usual $\{h_i\}$ distributed in $(-e, e)$ and performed a first CG-minimization at $\Gamma_{in} = 4$. Then, the transverse field Γ was repeatedly decreased in steps of $\Delta\Gamma \approx 0.2$, and at each step the new minimum was searched, via CG, starting from the minimum found in the previous step, down to the desired value of $\Gamma = 0.1$. This "variational annealing" procedure provides results for $\epsilon_{var}^{(MF)}$ which are distributed much closer to the classical ground state ϵ_{GS}, although we still obtain, in a non-negligible fraction of the attempts, quite bad minima. Such a lack of regularity, consequence of the multiplicity of metastable minima, makes $\psi_T^{(MF)}$ in Eq. 5.17 not a particularly suitable guiding wavefunction for a GFMC simulation.

Turning now to the Boltzmann-like choice in Eq. 5.20, the expectation value of the energy, as a function of the single parameter β, was calculated by a *Variational Monte Carlo* (VMC): in essence, an equilibrium GFMC simulation using $\psi_T^{(\beta)}$ as importance sampling trial wavefunction, but with all weights w_n consistently ignored, and taken to be $w_n = 1$ (see App. D.1, and references therein, for an introduction to VMC). [13] Fig. 5.2 shows (top panel) the optimal value β_{opt} of β which minimizes the variational energy $\langle\psi_T^{(\beta)}|H|\psi_T^{(\beta)}\rangle$, for several values of the transverse field Γ. [14] Notice that β_{opt} saturates for small Γ to about $\beta = 1$, somewhat surprisingly, since, for $\Gamma \to 0$, we would expect $\beta_{opt} \to +\infty$, so that the GS dominates. This is the effect of an ergodicity breaking of the algorithm. Indeed, from the classical spin-glass physics [29, 69] we know that a threshold energy E_{th} exist below which the system has a finite complexity, i.e., it displays an exponentially large (in N) number of metastable minima. Close to this threshold energy, the relaxation of any local algorithm toward equilibrium starts to be exceedingly slow, and the physical quantities measured are not really representative of their true averages. Evidently, for $\Gamma \to 0$, the variational algorithm

[12] Unfortunately, it is not at all obvious that an adiabatic theorem holds in this case, and there is therefore no guarantee that – even for an infinitely slow annealing – the optimal solution will be found.

[13] The averages are taken over 10^5 configurations of a single-walker Markov chain, separated by 10^3 Monte Carlo steps, in order to get uncorrelated data. The error is computed using the blocking technique [52]. We noticed that the correlation time, and the initial equilibration time, increase, as expected, when Γ is decreased inside the glassy low-Γ region. This is the main reason for using a single walker. Taking many walkers together would increase the statistics, but at the price of wasting many initial equilibration parts, which are expensive for small Γ. The correlation time, instead, it is not affect by the walker number in such a variational scheme.

[14] β_{opt} is found "by hand" via a cubic interpolation of results on a grid of β values.

is not visiting the regions near the true minima of the classical energy, but it is getting trapped in a band metastables states at higher energy. The central and bottom panels in Fig. 5.2 show the optimal variational energies $\epsilon_{var}^{(Boltz)} = \langle \psi_T^{(\beta_{opt})} | H | \psi_T^{(\beta_{opt})} \rangle / N$, and the variational residual energy $\epsilon_{res} = \langle \psi_T^{(\beta_{opt})} | H_{cl} | \psi_T^{(\beta_{opt})} \rangle - \epsilon_{GS}$ corresponding to the optimal β shown in the Top panel, for several values of transverse field Γ. For large Γ values, the variational total energy (center panel) is linear in Γ, as it should, since the transverse field kinetic term dominates in the quantum paramagnetic phase, while the variational residual energy per-site is of order 1. By decreasing Γ we notice that the variational residual energy saturates, for small Γ, to finite non-zero values, of order 0.03, in close agreement with the previously noted saturation in the optimal β, due to ergodicity breaking. Notice, however, that this saturation value is definitely below the best results provided by the previously discussed $\psi_T^{(MF)}$, of order 0.04, shown for comparison by a dashed horizontal line. Therefore, with all its pitfalls, the Boltzmann-like trial wavefunction in Eq. 5.20 provides quantitatively better residual energies, for small Γ, than the mean-field one in Eq. 5.17, and is also much better behaved as far as the minimization problem is concerned: for these reasons, we have decided to work out our GFMC results for the Boltzmann-like wavefunction only.

5.3 GFMC results I: Fixed Γ simulations

In this section we shall present the results obtained, for fixed values of Γ, by employing the whole GFMC machinery discussed in Sec. 5.1, with the Boltzmann-like trial wavefunction presented in Sec. 5.2. A pseudo-code description of the GFMC algorithm employed can be found in App. D.2.

In Fig. 5.3 we plot the GFMC results obtained for the 80×80 random Ising model instance used in Ref. [10], for several fixed values of the transverse field Γ. [15] The

[15] We list a few technical ingredients. We made use of 10 walkers, and we worked out the statistical average over 10^4 configurations separated by 10^4 Monte Carlo steps, skipping the first 10^3 steps for the equilibration. This statistics is reduced to only 10^3 configurations taken every 10^6 steps for $\Gamma \leq 0.25$. Branching is performed every GFMC step (Tests accomplished allowing for branching every 10 GFMC steps showed weight instability at small value of Γ (see also next section).). Indeed the small Γ correlation effects are more severe than in the variational case, and we were forced to run longer simulations in order to have comparable statistical errors. Moreover, one has two balance the bias effect due to a finite walker population – which increases the correlation – and the computational cost required by a multi-walker simulation. We also stress that for very small transverse fields the

top panel shows the GFMC estimate of the total energy per site for several values of Γ, compared to the variational results $\epsilon_{var}^{(Boltz)}$ illustrated in the previous section. The inset allows one to appreciate the differences between the two results in the small Γ region, which are invisible on the scale of the main plot. In the bottom panel we report several data regarding the residual diagonal energy. A small technical point is here in order. GFMC calculates directly the total energy estimates – $\langle H \rangle_w$ in the notation of App. D.1 –, while averages of operators that *do not commute* with the Hamiltonian H are less straightforwardly obtained. In particular, what one can simply evaluate is the so-called *mixed average* of the potential energy H_{cl},

$$\langle H_{cl} \rangle_w = \frac{\langle \psi_T | H_{cl} | \psi_{GS} \rangle}{\langle \psi_T | \psi_{GS} \rangle} ,$$

(see Sec. D.1 for more details). This is the estimator labeled "MIXED" in the second panel. The true expectation value we want is, instead, $\langle \psi_{GS} | H_{cl} | \psi_{GS} \rangle$, which might be poorly approximated by $\langle H_{cl} \rangle_w$ if the trial wavefunction is poor. A simple partial cure to this drawback, is to include the so-called *Ceperley correction* (see Eq. D.12):

$$\langle \psi_{GS} | H_{cl} | \psi_{GS} \rangle \approx 2 \langle H_{cl} \rangle_w - \langle H_{cl} \rangle_T , \tag{5.21}$$

where $\langle H_{cl} \rangle_T = \langle \psi_T | H_{cl} | \psi_T \rangle$ is the variational estimate. The results of the latter approximation for the residual energy are labeled "CEPERLEY" in Fig. 5.3, and are seen to be consistently lower than the mixed average estimates, which, in turn, are lower than the variational result. [16] However, the non-monotonic behavior of the

off-diagonal transition probabilities are so small that the internal time increases a lot between two reconfigurations, giving serious problems due to weight divergences (see Eqs. D.13 and D.14). This drawback will appear again in the GFMC annealing case (see Sec. 5.4). We compute the error bars using the blocking technique.

Usually a good indicator of stability of the simulation is the number of survived walkers after every reconfiguration (see App. D.1 for more details), which has to be at least $80\% - 90\%$. A smaller ratio yields larger bias effects due to finite population (see Sec. D.1), or – even worse – it is the fingerprint of a bad trial function. Whenever this happens, a smaller branching-time must be chosen (see App. D.2). A larger survival ratio ($\sim 100\%$) is instead not dangerous, but it often leads to very correlated data and, consequently, the need of longer simulations.

[16] Notice that this technical problem in the different GFMC estimates of the residual energy should not influence the $\Gamma \to 0$ region of the data, where the total energy and the potential energy approach each other, and the mixed average becomes, in principle, exact. As a consequence, we need not worry about this subtle differences in estimators in discussing the annealing results.

residual energy data for small $\Gamma < 0.25$, previously noted for the variational results, should ring a bell about the quality of the trial wavefunction, and the efficiency of the sampling, in that region.

5.4 GFMC Results II: Quantum Annealing.

In this section we shall finally present the results obtained by applying a GFMC-based QA to the 80×80 random Ising model instance considered in Ref. [10]. The basic algorithm illustrated in Sec. 5.1, and used for fixed Γ calculations in Sec. 5.3, is now used as an annealing tool, by reducing at each Monte Carlo step (MCS) the coupling Γ in the usual linear way, starting form an initial large value $\Gamma_0 = 2.5$, $\Gamma_n = \Gamma_0(1 - n/\tau)$. τ is here the annealing time measured as the total number of MCS performed by the algorithm, and we recall that, due to the continuous-time strategy, each MC step performs an *off-diagonal* move (flipping of a single spin), while the internal time of the algorithm advances according to the rate of escaping from a given configuration (see Eq. 5.16). For each value of Γ_n the trial wavefunction used is the Boltzmann-like one, defined in Eq. 5.20, with a variational parameter $\beta_{opt}(\Gamma_n)$, which corresponds to the instantaneous optimal value. [17] As for the branching, we employed $M = 20$ walkers, and we chose to perform a weight reconfiguration at each MCS. This choice is dictated by the fact that for small transverse fields Γ the the average weight is quite unstable; it would be definitely unnecessary for the initial stage of the annealing, where the weights are well under control (indeed, the walker branching survival rate is here $\sim 100\%$), but we decided to adopt a uniform choice all along the simulation. Even with this very conservative choice, the weights go completely wild if a Γ of order 10^{-7} is reached: for this reason, we have decided to cut-off the Γ_n annealing schedule in such a way that the final Γ is 10^{-4} and not 0. This guarantees a good weight stability. [18]

In Fig. 5.4 we report the final part of a GFMC-QA dynamics for a fixed annealing time $\tau = 10^8$. One can see (upper panel) that the average (over the walker population) of the total energy presents large fluctuations (spikes), which are due to the aforementioned weight instabilities, while the classical energy (lower panel) has a

[17] Practically, we have used for $\beta(\Gamma)$ the fitting function shown in Fig. 5.2, upper panel.

[18] Whenever it was possible to perform annealings with smaller cut-offs on Γ, and approximatively the same slope $\frac{1}{\tau}$, we checked that the results obtained are not very sensitive to the cut-off chosen.

more regular profile. Fig. 5.5 shows the best residual energy per spin ever reached during the annealing simulation, for several values of τ, averaged over 10 repetitions of the annealing (due to time-limitations, a single run is shown for the largest $\tau > 10^8$ annealings). The CA and PIMC-QA data obtained in Ref. [10] are also shown, for comparison. Notice that the τ axes of the three calculations are completely unrelated: the GFMC τ is measured in units in which a MCS is just a *single spin-flip*, while MCS for the CA and PIMC-QA are intended as *sweeps of the entire lattice* of N spins (including all the Trotter slices, for the PIMC case). For this reason, a comparison with the CA and PIMC-QA data with τ multiplied by a factor N is also given. Although the GFMC-QA data are strictly below both the CA and the PIMC-QA data, on the same per-spin time-unit (shifter CA and PIMC-QA data), it is clear that the GFMC-slope is definitely *worse* than that of PIMC-QA, and indeed surprisingly similar to CA. Moreover, the CPU-time needed for a single spin-flip in GFMC is much larger than the corresponding single spin move in CA or PIMC-QA (each GFMC move costs of order N operations), so that, at present, GFMC is not a real competitor of, say, PIMC-QA.

A few comments are in order. First of all, the small number of walkers used ($M = 20$), and the branching performed at every MCS might lead to a big bias of the energy estimators (see App. D.1). However, further tests where branching was performed every 10 MCS, provided only small improvements, not enough to close the gap to the PIMC-data slope.

The crucial point, we believe, has to do with the form and quality of the trial wavefunction used. The fact that importance sampling is indeed a necessary ingredient is confirmed by a series of GFMC annealing simulations (not shown) which we have performed with $\psi_T = 1$ (no importance sampling): the results clearly indicate that the system always gets stuck in the quantum paramagnetic phase (the classical energy per spin was around zero even for moderately long annelings, τ up to 10^8). Nevertheless, while a trial wavefunction is definitely needed, several signs – such as the quality of the fixed-Γ residual energy in Fig. 5.3, and the sizable differences between the various estimators there presented, as well as the noticed weight instabilities at low Γ – point toward the possibility that the Boltzmann-like ψ_T we have adopted is probably not the most suitable one. Perhaps, even the surprising coincidence of the GFMC-QA and CA data slopes is the result of such a choice for ψ_T, which might force or bias – in some sense – the instantaneous spin configurations to be distributed

according to the Boltzmann distribution.

So, the crucial question is: how to find a good trial variational wavefunction which describes well enough the small transverse field phase of the Ising spin-glass? A combination of the two different choices analyzed in Sec. 5.2, i.e., a Boltzmann-like factor supplemented by local fields h_i, is the first possibility, certainly worth investigating. [19] As a further possibility, one might introduce pair-correlations into the trial wavefunction – for instance by means of a spin-spin Jastrow factor, either at nearest-neighbor or at longer range – as usually done in the framework of correlated lattice models [71], and of electronic structure calculations [51, 72]. Unfortunately, for a spin-glass, due to frustration and disorder, the form of such pair-correlations is far from obvious.

In conclusion, we have identified the choice of the importance sampling wavefunction in a GFMC-QA as the crucial step, as well as a weak point, of the method. The results obtained are strongly influenced by such a choice: while no importance sampling ($\psi_T = 1$) produces useless results, a Boltzmann like choice $\psi_T \propto e^{-\beta^2 E_{cl}}$ leads to results with a slope very similar to CA – but computationally much more expensive! –, and definitely worse than PIMC-QA.

[19] The optimization of the large number of parameters, however, requires the use of more advanced minimization techniques, see [70], since CG is not possible for a correlated wavefunction.

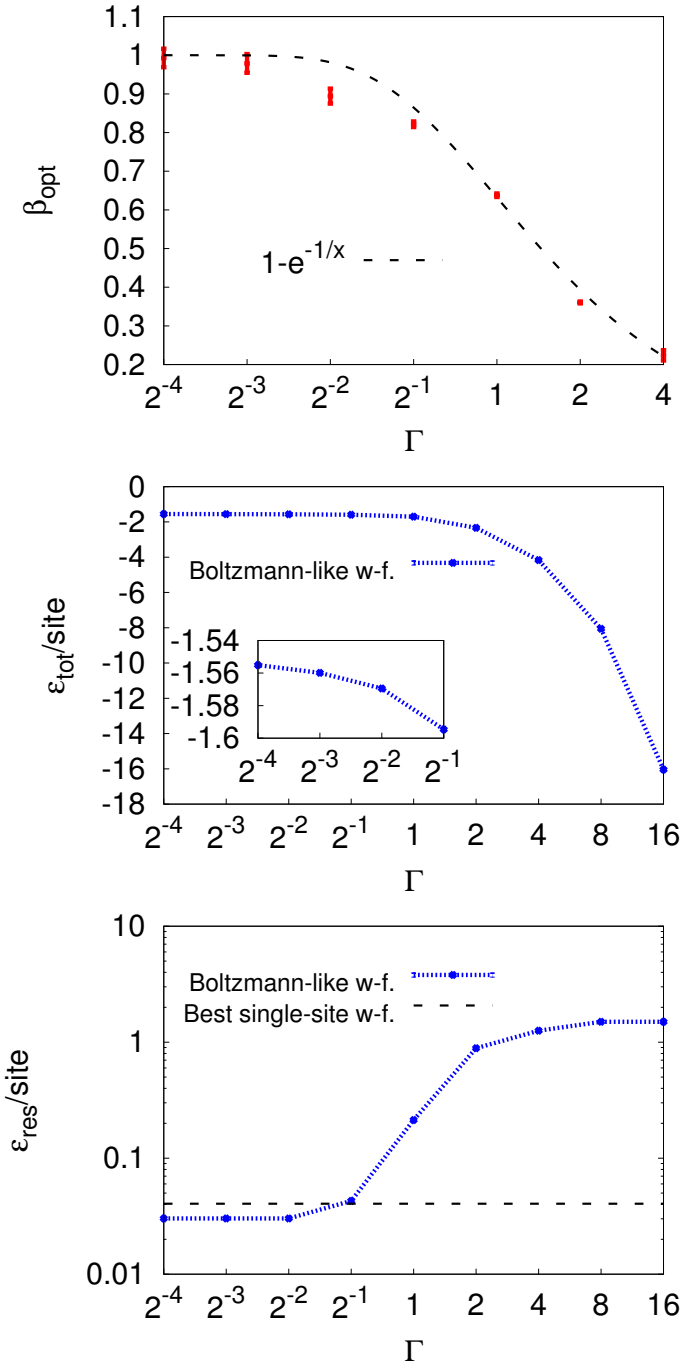

Figure 5.2: (Top) Plot of the optimal β, β_{opt}, for the trial wavefunction $|\psi_T^{(\beta)}\rangle$ defined in Eq. 5.20, for several value of the transverse field Γ. The dashed line is a fit of the data with the function $\beta_{opt}(\Gamma) = 1 - e^{-1/\Gamma}$. (Center) Optimal variational energies $\epsilon_{var}^{(Boltz)} = \langle\psi_T^{(\beta_{opt})}|H|\psi_T^{(\beta_{opt})}\rangle/N$ corresponding to the β_{opt} shown in the Top panel, for several Γ. The inset magnifies the small-Γ region. (Bottom) The variation residual diagonal energy $\epsilon_{res} = \langle\psi_T^{(\beta_{opt})}|H_{cl}|\psi_T^{(\beta_{opt})}\rangle - \epsilon_{GS}$ corresponding to the β_{opt} shown in the Top panel, for several Γ. The dashed horizontal line labeled 'MF' represents the best residual energy ever achieved, for $\Gamma > 0.01$, by employing the mean-field trial wavefunction in Eq. 5.18.

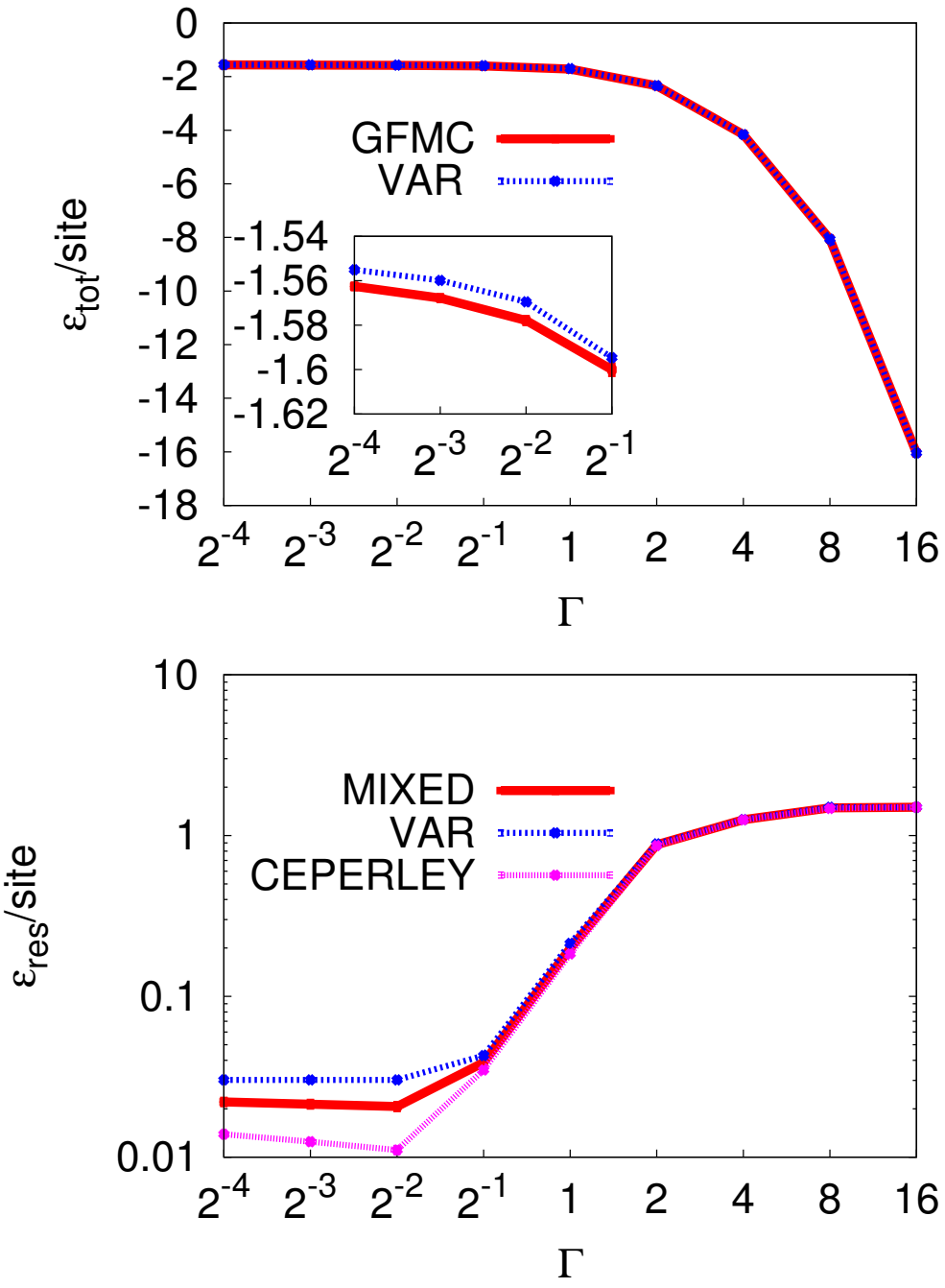

Figure 5.3: Results of the GFMC applied to the 80×80 random Ising model instance used in Ref. [10] for several fixed values of the transverse field Γ. For each Γ, we use as importance-sampling trial wavefunction the optimized $\psi_T^{(\beta_{opt})}$ discussed in Sec. 5.2. (Top) GFMC estimate of the total energy per spin $\langle H \rangle / N$, which is compared in the inset to the variational results of Fig. 5.2. (Bottom) Several GFMC estimators of the residual diagonal energy $\epsilon_{res} = \langle H_{cl} \rangle / N - \epsilon_{GS}$: the mixed average, the variational result of Fig. 5.2, and the Ceperley correction (see text and Eq. D.12)

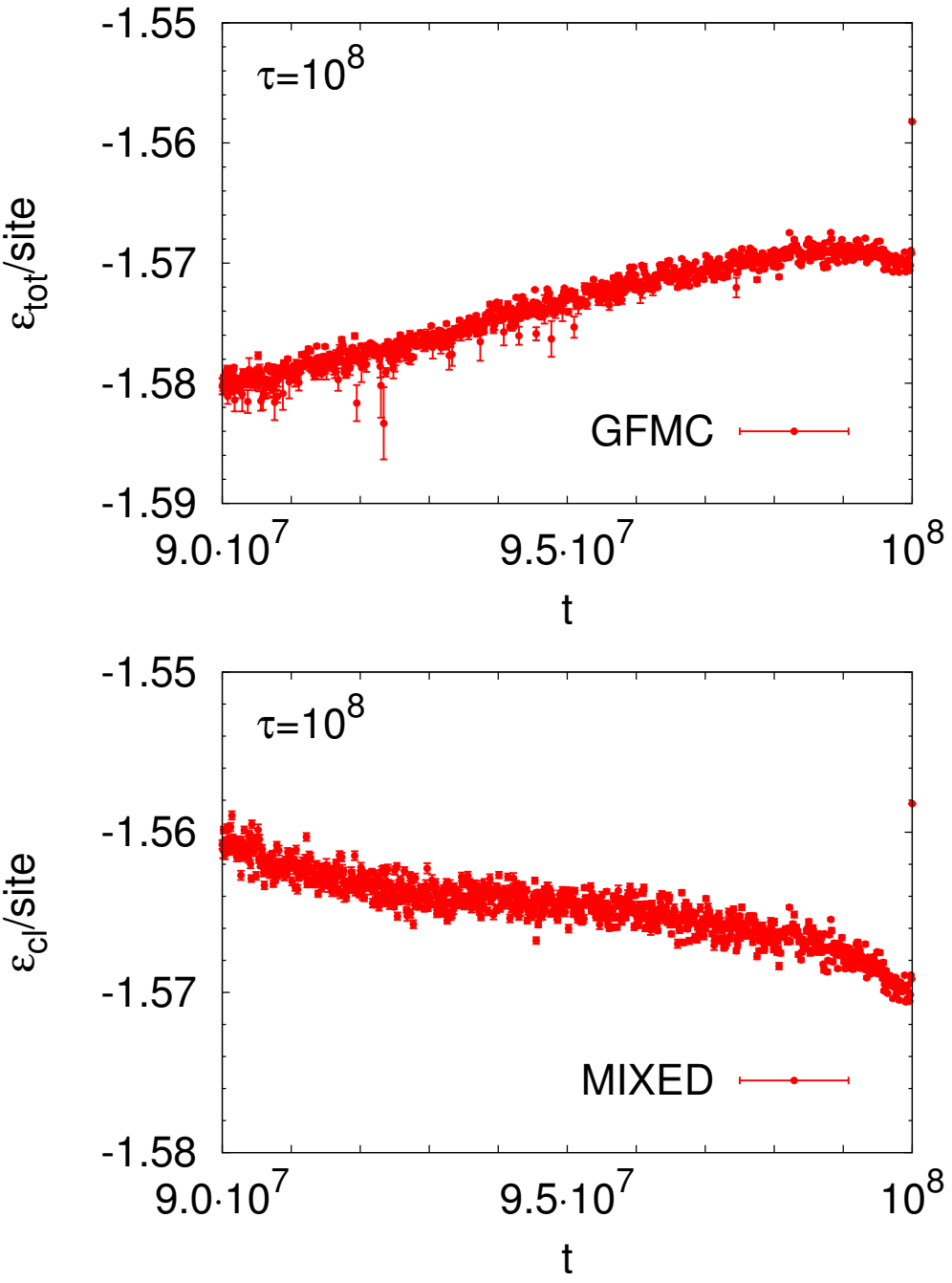

Figure 5.4: GFMC-QA dynamics for a fixed annealing time $\tau = 10^8$, in the time-interval $[0.9\tau, \tau]$. We employed the optimal trial function $|\psi_T^{(Boltz)}\rangle$, and $\Gamma_0 = 2.5$ as initial transverse field. (Top) GFMC estimate of the total energy per spin $\langle H\rangle/N$. (Bottom) MIXED estimators of the classical energy $\epsilon_{cl} = \langle H_{cl}\rangle/N$.

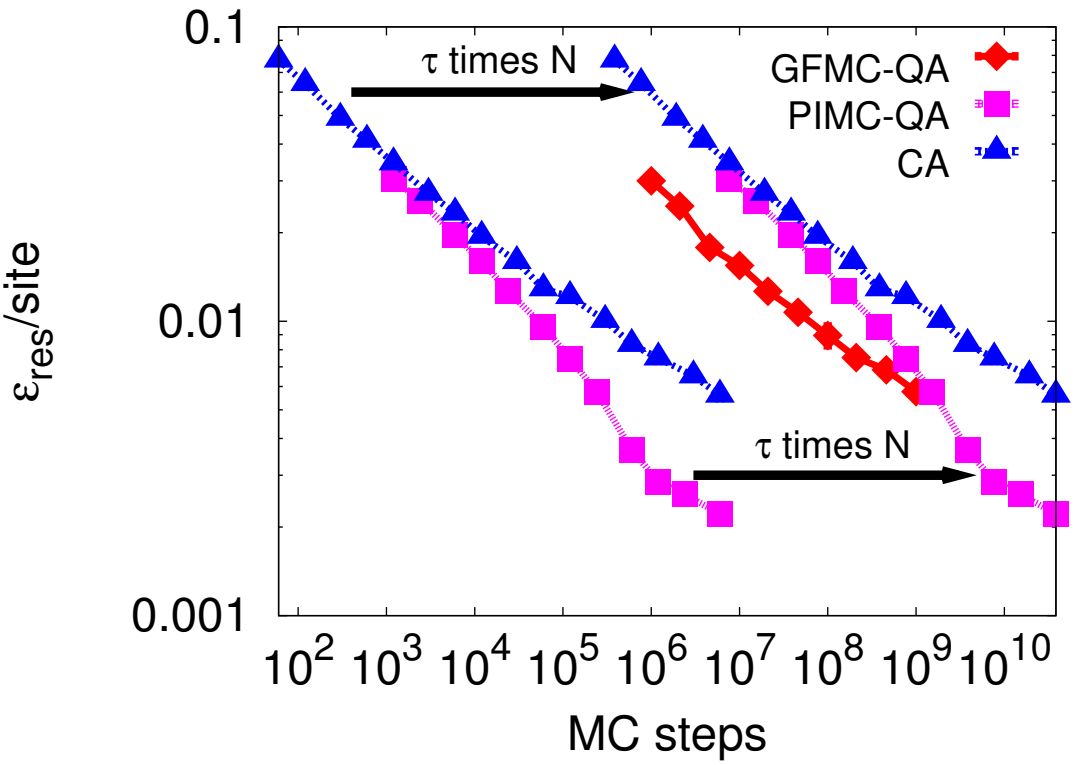

Figure 5.5: The average best residual energy obtained by GFMC-QA for the 80×80 instance of the random Ising model studied in Refs.[10, 11], versus the total annealing time τ. Previous results obtained by CA and PIMC-QA with $P = 20$ Trotter slices [10, 11] are shown for comparison. The GFMC time-unit is a single spin-flip, while CA and PIMC-QA time units are sweeps of the entire lattice (see Ref. [11]). Importance sampling is performed by using the optimal trial wavefunction $|\psi_T^{(Boltz)}\rangle$ of Sec. 5.2. The transverse field is linearly reduced down to 10^{-4} in a total annealing time τ, starting from $\Gamma_0 = 2.5$ (the GFMC time-step consists of a single spin-flip).

Conclusions and Perspectives

In this thesis I have addressed some important issues about classical and quantum annealing (QA). I have illustrated several applications of both annealing strategies to a range of problems going from textbook toy-models, to a challenging hard optimization problem, the random Ising model, using techniques which go from deterministic Fokker-Planck or Schrödinger evolutions, for the toy models, to a novel implementation of a Green's Function Monte Carlo QA scheme, for the Ising case.

I started with a very simple toy-model, where the energy landscape to be optimized was well under control: a simple double-well potential in one dimension. There I discovered the crucial difference between a classical annealing (CA) approach, realized by the solution of a time-dependent Fokker-Planck's (FP) equation, and a QA dynamics, where the Schrödinger equation (either in real or in imaginary time) governs the system. In the FP classical case, the system reduces, after a very short annealing, to an effective *discrete two-level system*, whose dynamics is governed by the ratio between the level splitting Δ_V and the barrier height B, with a residual energy $\epsilon_{res}(\tau) \sim \tau^{-\Delta_V/B}$. In the Schrödinger case, on the contrary, the spectrum of the system is never really a pure two-level system one, and the relevant feature controlling the dynamics is a possible Landau-Zener avoided crossing in the ground state related to the tunneling amplitude through the barrier. The analysis of other simple toy models gave us other interesting observations. For instance, while a series of high barriers without disorder, as in a parabolic washboard potential, is enough to make CA logarithmically slow, QA turned out to be much more effective in that case, due to a particular simplicity of the Schrödinger's eigenvalue spectrum. As a by-product of our study, I have come to the conjecture, verified by all the cases analyzed, that doing real-time or imaginary-time Schrödinger dynamics is essentially equivalent, the latter being quantitatively better: this opens up the route to all kind of stochastic implementations of QA, notably more suited to the imaginary-time case. Finally, the

crucial role of disorder in making QA (as well as CA) slow is discussed in one of the simplest context: the one-dimensional disordered Ising ferromagnet, which has a trivial ferromagnetic ground state, but whose annealing dynamics is made slow by the presence of possibly large domains of wrongly aligned spins, pinned by arbitrarily weak J_i. An explicit study of the 1D disorder Ising case would be a very instructive future work in this line of approach.

Next, armed with the understanding gained by the deterministic approaches, I studied the same simple double-well potential through Monte Carlo (MC) annealing approaches, both in the classical and in the quantum case. In the classical case, I discovered the large freedom that the choice of the proposal move gives to the method, with results that differ very much if one consider a Box, a Gaussian, or a Lorentzian distribution. This ambiguity of the MC-CA outcome was rationalized through a detailed numerical and analytical study of the spectral properties of the associated Markov transition matrix W, which also provided a key to the optimization of the move range σ during the annealing, through the search of the instantaneous maximum gap of W. In the quantum case, I discovered how an allegedly *state-of-the-art* Path-Integral Monte Carlo (PIMC) based QA, with smart bisection moves and improved fourth-order action, can perform in a disappointing poor way in the presence of a Landau-Zener tunneling event: the rather *ad-hoc* solution to the problem – adding *instanton* moves to the sampling – is playing an unfair game, since it exploits a knowledge of the landscape which is generally not available. Finally, the crucial role played by the choice of the kinetic energy operator was simply tested, in the PIMC-QA double-well case, by comparing the familiar non-relativistic Laplacian term (associated to a Gaussian propagator) with a *relativistic* dispersion $H_{kin} = \Gamma|p|$ (associated to a Lorentzian propagator), which turns out to perform much better than the former. In some sense, the *Leitmotiv* of these MC studies is that the choice of moves (in the classical case), and the choice of moves plus the form of H_{kin} (in the quantum case), determine in some way the *effective energy landscape* of the problem, well beyond the pure potential energy term.

As a final application of QA, I decided to test a Green's Function Monte Carlo (GFMC) based QA on a real optimization problem, the random Ising model, where both CA and PIMC-QA results are already known. In principle, GFMC looked like a promising tool. In particular, at variance with PIMC, GFMC does not suffer: i) from finite temperature effects, which give a thermal lower bound to the residual energies

attained; ii) from difficulties in the Trotter discretization, as noticed in the Traveling Salesman Problem case (see Sec. 1.2); iii) from difficulties in the sampling of the action, as notice in the double-well case. However, GFMC needs, as an essential ingredient, a good variational *Ansatz* for the ground state, to be used as importance-sampling trial wavefunction. This need for a good trial wavefunction turns out to be a crucial and difficult step in the GFMC-QA approach: the two wavefunctions I have tested, a mean-field single-site product and a Boltzmann-like *Ansatz*, both provide evidently poor choices. The final GFMC-QA results obtained using the Boltzmann-like trial wavefunction turn out to be definitely worse than PIMC-QA, with a slope which is very similar to the CA data. Improvements in the trial-wavefunctions are, evidently, an open issue for future work in this field.

In conclusion, the crucial question "*Is Quantum really better than Classical?*" has now a more clear answer. Although in many of the examples illustrated QA often wins over CA, sometimes it does not, like in the 3-SAT case (see Ref [16]). The final results is *a priori* not guaranteed and the outcome of the battle is strongly related to the (effective) energy landscape of the problem one deals with, which is in turn influenced by the choice of Monte Carlo dynamics, and by the quantum kinetic energy used. Moreover, it is quite clear that QA, although potentially useful and sometimes more convenient than CA, is not capable, in general, of finding solutions of NP-complete problems in polynomial time. Nevertheless, understanding when and how quantum mechanics can quantitatively improve on the solution of hard optimization problems is still a timely issue, which deserves further studies.

Appendix A

Appendix of Chapter 2

A.1 Annealing in a parabolic potential

A.1.1 Classical Annealing (Fokker-Planck) case

It is straightforward to find the solution of the Fokker-Planck Eq. 2.2 when the the potential is harmonic, $V(x) = kx^2/2$, and the initial condition is the Boltzmann distribution $P(x, t = 0) \propto \exp(-kx^2/(2k_BT_0))$. Indeed, it is simple to verify that the Gaussian *Ansatz*

$$P(x,t) = C_t e^{-B_t x^2} , \tag{A.1}$$

fulfills the initial condition (if $B_{t=0} = B_0 = k/(2k_BT_0)$) and solves the FP equation, as long as the two functions B_t and C_t satisfy the following ordinary differential equations:

$$\begin{cases} \dot{B}_t = 2D_t \left(\frac{kB_t}{k_BT(t)} - 2B_t^2 \right) \\ \dot{C}_t/C_t = D_t \left(\frac{k}{k_BT(t)} - 2B_t \right) \end{cases} . \tag{A.2}$$

The normalization constant C_t turns out to be irrelevant in calculating the average potential energy which we need

$$\epsilon_{pot}(t) = \frac{\int dx V(x) P(x,t)}{\int dx P(x,t)} = \frac{k}{4B_t} , \tag{A.3}$$

and can be completely forgotten, since the equation for B_t does not involve it. The equation for B_t appears to be *non-linear*, but can immediately transformed into a linear equation by dividing up both sides by B_t^2 and recognizing that the correct variable to use is precisely $1/B_t$, or better yet, $\epsilon_{pot}(t)$. In terms of $\epsilon_{pot}(t)$ we can therefore write a linear equation of the form:

$$\frac{d}{dt}\epsilon_{pot}(t) = kD_t \left(1 - \frac{2}{k_BT(t)}\epsilon_{pot}(t) \right) , \tag{A.4}$$

the initial condition being simply given by the equipartition value $\epsilon_{pot}(t = 0) = (k/4B_0) = k_BT_0/2$. An alternative way [42, 43] of deriving Eq. A.4 consists in taking the derivative with respect to time

of both sides of Eq. A.3, using then the FP equation for $\partial P/\partial t$ on the left hand-side of the ensuing equation, and finally integrating by parts the terms containing spatial derivatives of P (this procedure results in a closed differential equation for $\epsilon_{pot}(t)$ only if the potential $V(x)$ is harmonic). As every one-dimensional linear first-order differential equation, Eq. A.4 can be integrated by quadrature, the solution being:

$$\begin{aligned}\epsilon_{pot}(t) &= \epsilon_{pot}(t=0)\, e^{-(k/2k_B)\int_0^t dy D_y/T(y)} \\ &\quad + k\int_0^t dt' D_{t'} e^{-(k/2k_B)\int_{t'}^t dy D_y/T(y)} \,.\end{aligned}$$

If we now anneal the temperature down to 0 in a time τ in the usual way, $T(t) = T_0(1-t/\tau)^{\alpha_T}$, assuming that the diffusion constant behaves as $D(T) = D_0(T/T_0)^{\alpha_D}$, we readily get an analytic expression for the residual energy at the end of the annealing, $\epsilon_{res}(\tau) = \epsilon_{pot}(t=\tau)$, which can be shown to behave, for large τ, as a power law:

$$\epsilon_{res}(\tau) \approx \tau^{-\Omega_{CA}} \qquad \Omega_{CA} = \frac{\alpha_T}{\alpha_T(\alpha_D - 1) + 1} \,. \tag{A.5}$$

Quite evidently, annealing proceeds here extremely fast, with a power-law exponent Ω_{CA} that can increase without bounds (for instance if $\alpha_D = 1$) upon increasing the exponent α_T of the annealing schedule. Notice, however, that large values of α_D are, on the contrary, detrimental.

A.1.2 Quantum Annealing (Schrödinger) case

Consider now the problem of a particle moving in a parabolic potential $V(x) = kx^2/2$, with a time-dependent mass, such that the Hamiltonian is given by:

$$H(t) = \frac{k}{2}x^2 - \Gamma(t)\nabla^2 \,,$$

where $\Gamma(t) = \hbar^2/2m(t)$ denotes the coefficient of the Laplacian operator. The Schrödinger evolution of the wavefunction $\psi(x,t)$ is then,

$$\xi\partial_t\psi(x,t) = H(t)\psi(x,t) \,, \tag{A.6}$$

where $\xi = i\hbar$ for a real time (RT) evolution, and $\xi = -\hbar$ for an imaginary time (IT) evolution. Now, whereas general solutions of the time-dependent Schrödinger equation for arbitrary initial condition $\psi(x, t=0)$ are not easy, it turns out that if $V(x)$ is quadratic, then any initial Gaussian wavefunction propagates into a Gaussian, which is enough for our goal. In detail, write the following *Ansatz* for $\psi(x,t)$:

$$\psi(x,t) = C_t e^{-B_t x^2/2} \qquad \mathrm{Real}(B_t) = \Re(B_t) > 0 \,. \tag{A.7}$$

Substituting the *Ansatz* for $\psi(x,t)$ into the Schrödinger evolution (in RT or in IT), one immediately verifies that $\psi(x,t)$ satisfies Eq. A.6 for arbitrary $\Gamma(t)$ as long as B_t and C_t satisfy the following ordinary differential equations:

$$\begin{cases} -\xi\dot{B}_t = k - 2\Gamma(t)B_t^2 \\ -\xi\dot{C}_t/C_t = -\Gamma(t)B_t \end{cases} \,. \tag{A.8}$$

Once again, the normalization constant C_t turns out to be irrelevant in calculating the average potential energy

$$\epsilon_{pot}(t) = \frac{\int dx V(x)|\psi(x,t)|^2}{\int dx |\psi(x,t)|^2} = \frac{k}{4\Re(B_t)} ,$$

and can be completely forgotten, since the equation for B_t does not involve it. The initial condition $B_{t=0}$ is, by assumption, such that at $t = 0$ the system is in the *ground state* corresponding to $\Gamma_0 = \Gamma(t = 0)$. Such a ground state value B_0 is easily calculated by equating to zero the right-hand side of Eq. A.8 with $\Gamma(t) \to \Gamma_0$, i.e., $B_0 = \sqrt{k/(2\Gamma_0)}$.

A few general considerations can be based on a purely qualitative analysis of Eq. A.8. Consider first the IT case, where the equation for B_t reads $\hbar\dot{B}_t = k - 2\Gamma(t)B_t^2$. One can easily get convinced that B_t is forced to be a real, positive and monotonically increasing function of t, i.e., $B_t > 0$ and $\dot{B}_t > 0$ for $t > 0$. Therefore, we can easily write an inequality of the form:

$$\hbar\dot{B}_t = k - 2\Gamma(t)B_t^2 \leq k ,$$

from which we immediately conclude, by integrating over t the two sides of the inequality, that

$$B_\tau - B_0 \leq \frac{k\tau}{\hbar} ,$$

i.e., the residual energy $\epsilon_{res}(\tau) = k/(4B_\tau)$ cannot decrease faster than $1/\tau$ for $\tau \to \infty$.

We will now assume, without great loss of generality, that the Laplacian coefficient $\Gamma(t)$ is given by $\Gamma(t) = \Gamma_0 f(t/\tau)$, where τ is an annealing time-scale (for instance the annealing time, when a linear schedule is used) and $f(t')$ is a positive decreasing function for $t' \geq 0$ such that $f(t' \leq 0) = 1$. It is useful to switch to dimensionless variables by measuring times in unit of τ, $t' = t/\tau$, and B_t in units of its initial ground state value $B_{t=0} = B_0$. The appropriate dimensionless quantity is therefore $b(t';\tau) = B_t/B_0$, with $t' = t/\tau$, where the parametric dependence on the annealing time-scale τ has been explicitly indicated. The equation for $b(t';\tau)$ is given by:

$$\begin{aligned} \dot{b}(t';\tau) &= \alpha[1 - f(t')b^2(t';\tau)] \qquad \alpha = \tau\sqrt{2k\Gamma_0}/(-\xi) \\ b(0;\tau) &= 1 , \end{aligned} \tag{A.9}$$

where the dot, from now on, will denote a derivative with respect to t'. Notice that the parametric dependence on τ is all buried into the coefficient α, which reabsorbs also the $-\xi$ appearing in the dynamics (RT versus IT). This kind of non-linear differential equation is of the well known *Riccati* form. It can be transformed into a *linear second-order* differential equation by operating the following substitution

$$b(t';\tau) = \frac{\dot{y}(t';\tau)}{\alpha f(t')y(t';\tau)} , \tag{A.10}$$

where, evidently, y is defined up to an overall normalization constant. Indeed, simple algebra shows that we can re-express Eq. A.9 as a second-order linear equation for y, as follows:

$$\begin{aligned} &f(t')\ddot{y}(t';\tau) - \dot{f}(t')\dot{y}(t';\tau) = \alpha^2 f^2(t')y(t';\tau) \\ &\dot{y}(0;\tau) = \alpha y(0;\tau) . \end{aligned} \tag{A.11}$$

As long as $f(t') \neq 0$, it is simple to verify that the second-order equation for $y(t';\tau)$ can be also equivalently written as:

$$\frac{d}{dt'}\left(\frac{\dot{y}(t';\tau)}{f(t')}\right) = \alpha^2 y(t';\tau) \; . \tag{A.12}$$

Finally, denoting by $Y(t';\tau)$ the indefinite integral of $y(t';\tau)$, such that $\dot{Y} = y$, and integrating over t' both sides of Eq. A.12, we can also write:

$$\ddot{Y}(t';\tau) = \alpha^2 f(t') Y(t';\tau) \; . \tag{A.13}$$

Eq. A.13 is easily solved when the annealing schedule for $\Gamma(t)$ is linear, $\Gamma(t) = \Gamma_0(1 - t/\tau)$, i.e., when $f(t') = (1-t')^{\alpha_\Gamma}$ with $\alpha_\Gamma = 1$, a case in which Eq. A.13 is of the *Airy* type. In the latter case, it is useful to perform a final change of independent variable to $z = \xi^2|\alpha|^{\frac{2}{3}}(1-t')$, so that, defining $F(z) = Y(t';\tau)$, we can write the equation for F in the standard Airy form:

$$\frac{d^2}{dz^2}F(z) = zF(z) \; . \tag{A.14}$$

The general solution of Eq. A.14 is given in terms of the two Airy's functions $Ai(z)$ and $Bi(z)$,

$$F(z) = \beta Ai(z) + \gamma Bi(z) \; ,$$

where β and γ are two constant coefficients. Going back to $Y(t';\tau)$ and $y(t';\tau)$, we then have the explicit expressions:

$$\begin{aligned}
Y(t';\tau) &= \beta Ai(\xi^2|\alpha|^{\frac{2}{3}}(1-t')) + \gamma Bi(\xi^2|\alpha|^{\frac{2}{3}}(1-t')) \\
\dot{Y}(t';\tau) &= y(t';\tau) = -\beta\xi^2|\alpha|^{\frac{2}{3}} Ai'(\xi^2|\alpha|^{\frac{2}{3}}(1-t')) - \gamma\xi^2|\alpha|^{\frac{2}{3}} Bi'(\xi^2|\alpha|^{\frac{2}{3}}(1-t')) \\
\ddot{Y}(t';\tau) &= \dot{y}(t';\tau) = \beta|\alpha|^{\frac{4}{3}} Ai''(\xi^2|\alpha|^{\frac{2}{3}}(1-t')) + \gamma|\alpha|^{\frac{4}{3}} Bi''(\xi^2|\alpha|^{\frac{2}{3}}(1-t')) \\
&= \alpha^2(1-t')\left\{\beta Ai(\xi^2|\alpha|^{\frac{2}{3}}(1-t')) + \gamma Bi(\xi^2|\alpha|^{\frac{2}{3}}(1-t'))\right\} \; ,
\end{aligned} \tag{A.15}$$

where the prime indicates a derivative with respect to z, and we have used the property of Airy's functions that $Ai''(z) = zAi(z)$ and $Bi''(z) = zBi(z)$. Finally, substituting back the expressions in Eq. A.15 into the original function B_t, see Eq. A.10, we get:

$$B_t = B_0 \frac{\dot{y}(t/\tau;\tau)}{\alpha(1-t/\tau)y(t/\tau;\tau)} = \xi\, B_0 |\alpha|^{\frac{1}{3}} \frac{\beta Ai(\xi^2|\alpha|^{\frac{2}{3}}(1-t/\tau)) + \gamma Bi(\xi^2|\alpha|^{\frac{2}{3}}(1-t/\tau))}{\beta Ai'(\xi^2|\alpha|^{\frac{2}{3}}(1-t/\tau)) + \gamma Bi'(\xi^2|\alpha|^{\frac{2}{3}}(1-t/\tau))} \; .$$

This general solution correctly depends on one parameter only, i.e., γ/β, so that we can put $\beta = 1$ without loss of generality. We impose the initial condition $B_{t=0} = B_0$ by requiring:

$$\xi\,|\alpha|^{\frac{1}{3}} \frac{Ai(\xi^2|\alpha|^{\frac{2}{3}}) + \gamma Bi(\xi^2|\alpha|^{\frac{2}{3}})}{Ai'(\xi^2|\alpha|^{\frac{2}{3}}) + \gamma Bi'(\xi^2|\alpha|^{\frac{2}{3}})} = 1 \; .$$

Solving for γ, we get:

$$\gamma = -\frac{\xi^2|\alpha|^{\frac{1}{3}} Ai(\xi^2|\alpha|^{\frac{2}{3}}) - Ai'(\xi^2|\alpha|^{\frac{2}{3}})}{\xi^2|\alpha|^{\frac{1}{3}} Bi(\xi^2|\alpha|^{\frac{2}{3}}) - Bi'(\xi^2|\alpha|^{\frac{2}{3}})}$$

Due to the asymptotic properties of the Airy's functions ($Ai(z) \to 0$, $Ai'(z) \to 0$ and $Bi(z) \to \infty$, $Bi'(z) \to \infty$, when $z \to \infty$ on the real axis), we conclude that for the IT case:

$$\gamma \to 0 \qquad \text{for } \alpha \ \ (\text{or } \tau) \to \infty$$

so that, finally,

$$\begin{aligned} B(\tau) &= -B_0 |\alpha|^{\frac{1}{3}} \frac{Ai(0)) + \gamma Bi(0)}{Ai'(0) + \gamma Bi'(0)} \\ &\simeq B_0 \left(\frac{2k\Gamma_0 \tau}{\hbar} \right)^{\frac{1}{3}} \frac{\Gamma(\frac{1}{3})}{3^{\frac{1}{3}}\Gamma(\frac{2}{3})} , \end{aligned}$$

where we used that $Ai(0) = 3^{-2/3}/\Gamma(2/3)$ and $Ai'(0) = -3^{-1/3}/\Gamma(1/3)$. On the other hand, for the RT case we have to take the limit $z \to -\infty$ (on the real axis) instead, and in that region all the Airy's functions oscillate. Nevertheless is possible to show that the value of γ is uniformly bounded for $\tau \to \infty$.

We conclude, therefore, that for large τ and with a linear annealing schedule, $\alpha_\Gamma = 1$,

$$B(\tau) \propto \tau^{\frac{1}{3}} \tag{A.16}$$

and, consequently, the residual energy behaves asymptotically as

$$\epsilon_{res}(\tau) = \frac{k}{4\Re(B_\tau)} \propto \tau^{-\frac{1}{3}} . \tag{A.17}$$

The generalization of this result to an arbitrary annealing exponent $\alpha_\Gamma \geq 0$ in $\Gamma(t) = \Gamma_0(1 - t/\tau)^{\alpha_\Gamma}$, is a bit more involved. It is however possible to establish a generalization of Eq. A.17, for arbitrary α_Γ, that we checked by means of direct numerical integration. It reads:

$$\epsilon_{res}(\tau) = \frac{k}{4\Re(B_\tau)} \propto \tau^{-\Omega_{QA}} \qquad \Omega_{QA} = \frac{\alpha_\Gamma}{\alpha_\Gamma + 2} , \tag{A.18}$$

an expression that holds true for both RT and IT annealing.

A.2 Classical annealing with quantum tools: Imaginary time Schrödinger evolution of the Fokker-Planck equation

A side issue, but nonetheless an interesting one we wish to discuss here relates to classical annealing, and concerns the well-known mapping of a Fokker-Planck equation onto an imaginary-time Schrödinger problem [73], and its implications on the relationship between CA and QA. The bottom-line will be that the mapping does *not* imply that a FP-based CA is actually equivalent to QA, and moreover that such a mapping is not particularly useful in our annealing context.

Consider, once again, the FP problem with a time-dependent temperature $T(t)$

$$\frac{\partial}{\partial t}P(x,t) \;=\; \frac{1}{\eta_t}\,\mathrm{div}\,(P\nabla V) \;+\; D_t\nabla^2 P\;. \tag{A.19}$$

where both the friction coefficient and the diffusion constant are now time-dependent quantities, which we indicate by $\eta_t = \eta(T(t))$ and $D_t = D(T(t))$. In order to map the problem in Eq. A.19 onto an imaginary-time Schrödinger problem, the standard procedure [73] is to pose $P(x,t) = \Phi_0(x,t)\psi(x,t)$ and to determine $\Phi_0(x,t)$ in such a way as to eliminate the non-Schrödinger-looking drift term, turning it onto a standard potential term. The algebra is trivial. One can show that the drift term is eliminated if, and only if, the Φ_0 satisfies the equation:

$$\nabla\Phi_0(x,t) = -\frac{\nabla V}{2\eta_t D_t}\Phi_0(x,t)\;,$$

whose solution is readily found to be:

$$\Phi_0(x,t) = C(t)e^{-V/(2\eta_t D_t)}\;,$$

with $C(t)$ a function of time only, which can even be taken to be constant without loss of generality. By plugging $P = \Phi_0\psi$ in the FP Eq. A.19, with Φ_0 as above, one can show that the resulting equation for $\psi(x,t)$ is indeed of the Schrödinger form

$$-\frac{\partial}{\partial t}\psi(x,t) = -D_t\nabla^2\psi(x,t) + V_{FP}(x,t)\psi(x,t)\;, \tag{A.20}$$

with an effective potential V_{FP} given by

$$V_{FP}(x,t) = \frac{1}{2\eta_t}\left[\frac{(\nabla V)^2}{2\eta_t D_t} - \nabla^2 V\right] + \frac{\partial_t\Phi_0(x,t)}{\Phi_0(x,t)}\;.$$

The first term in V_{FP} is the standard effective potential of the Riccati form obtained in the time-independent case [73]. The second piece in V_{FP} is absent in the time-independent case, and can be easily seen to be [1]

$$\frac{\partial_t\Phi_0(x,t)}{\Phi_0(x,t)} = -V(x)\frac{d}{dt}\left(\frac{1}{2\eta_t D_t}\right)\;.$$

The main point we want to stress is that, by annealing $T(t)$ and hence D_t, we not only reduce the coefficient of the Laplacian in Eq. A.20, but we also strongly modify the *potential* V_{FP}, at variance with a genuine QA where only the kinetic term is annealed down. The modifications of the potential are so strong that, at low temperature, the instantaneous eigenvalue spectrum associated to the FP equation, as discussed in Sec. 2.2, is vastly different from that of the quantum double well system.

A.3 Technical remarks on discretized dynamics

In order to simulate the exact dynamics provided by a partial differential equation (like the Fokker-Planck or the Schrödinger equations), we first restrict the phase space to a compact interval of the

[1] If we consider that $\eta_t D_t = k_B T(t)$, we clearly see that, when annealing the temperature, the overall sign of the factor in front of $V(x)$ in the extra term of V_{FP} is negative.

real line and further we discretize it by means of an uniform grid of N_g points, whose mesh is a. As a consequence, the partial differential equation becomes a set of N_g ordinary differential equations: $\dot{x} = f(t, x)$.

A fast and accurate integration method for such a system is provided by the *fourth-order Runge-Kutta* scheme [74][page 710ff]. According to this method, if the system lies in a configuration x_n at time t_n, then at time $t_{n+1} = t_n + \Delta t$ it will stay in:

$$x_{n+1} = x_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6} + 0\left(\Delta t^4\right) \ ,$$

where the constants are given by:

$$\begin{aligned} k_1 &= \Delta t\, f(t_n, x_n)\,, \\ k_2 &= \Delta t\, f(t_n + \frac{\Delta t}{2}, x_n + \frac{k_1}{2})\,, \\ k_3 &= \Delta t\, f(t_n + \frac{\Delta t}{2}, x_n + \frac{k_2}{2})\,, \\ k_4 &= \Delta t\, f(t_n + \Delta t, x_n + k_3)\ . \end{aligned}$$

We notice that this algorithm needs four evaluations of the derivative operator $f(t, x)$. Moreover, in order to improve the precision of the algorithm, we properly employ the adaptive form of this method, which can be found in Ref. [74][page 714ff].

A little concern arises when – as like in the case of the Fokker-Planck and the Schrödinger equations – the operator $f(t, x)$ is linear. In particular $f(t, x)$ can be seen as a matrix and – since very deep and wide results are available about the iteration of a matrix in a linear space (see [49][page 53ff]) – one can analytically study the stability of a given algorithm.

The (disretized) Schrödinger equation is computational "safe", because the discretized Hamiltonian is symmetric:

$$(H\psi)_i = \frac{1}{2\,m}\,\frac{\psi_{i-1} - 2\,\psi_i + \psi_{i+1}}{a^2} + V_i\,\psi_i\ .$$

As a consequence, the exponential operator $e^{-i\,\Delta t\,H}$ is well defined for every value of Δt, so that the discrete evolution (with time step Δt) is stable. The imaginary evolution – which has been introduced in Chap. 2 – is also well behaved, since it is implemented by an evolution operator, $e^{-\,\Delta t\,H}$, which has the same good properties of $e^{-i\,\Delta t\,H}$.

Unfortunately this is no longer true for the (discretized) Fokker-Plank equation, that reads:

$$\begin{aligned} \frac{\partial\, P_i}{\partial\, t} &= (W\,P)_i = D\,\frac{P_{i-1} - 2\,P_i + P_i}{a^2} \\ &+ \frac{1}{\eta\,a^2}\left(P_{i+1}\,(V_{i+1} - V_i) + P_i\,(V_{i-1} - 2\,V_i + V_{i+1}) + P_i\,(V_{i-1} - V_i)\right)\ , \end{aligned}$$

because it is not symmetric. [2]

Anyway, the exponential of the operator W is stable and it converges to a unique equilibrium vector P_{eq} if the following two conditions hold: First, all the non-diagonal entries of $W_{i,j}$ must be

[2] Whether the reader is not yet acquainted with the Fokker-Planck equation, we refer to Sec. 2.1 for a brief introduction to our notation. Here we stress only few formal proprieties of its discretized form.

non-positive, secondary W must be *column normalized*: $\sum_{i=1}^{N_g} W_{i,j} = 1 \, \forall j$. The last constraint can be easily satisfied, taking care of the boundary conditions. The former is more delicate. In particular, it is equivalent to the condition: $D \geq \frac{a}{2} \left| \frac{\partial V}{\partial x} \right|_i \forall i$, so that, in order to have a stable algorithm, the discretization mesh must be small enough. Or equivalently, given a mesh, there is a lower limit to the value of the coefficient D that one can safely use during the simulations. This point matters in the case of the Classical Annealing, since D is indeed the annealing parameter (see Sec. 2.1) and it is usually decreased to zero.

Appendix B

Appendix of Chapter 3

B.1 Eigenvalues of a Fokker-Planck operator

Before approaching the general problem posed in Sec. 3.9, we discuss the much simpler case in which $\tilde{W}$ is a *local differential operator*, of the Fokker-Planck form (see Sec.2.1), i.e., the Master equation reads:

$$\frac{\partial}{\partial t} P(x,t) = -\bar{L}_{FP} P(x,t) \,, \tag{B.1}$$

where $\bar{L}_{FP}$ is the (non-Hermitian) Fokker-Planck operator

$$\bar{L}_{FP} = -D \frac{\partial^2}{\partial x^2} - \frac{1}{\eta} \frac{\partial V}{\partial x} \frac{\partial}{\partial x} - \frac{1}{\eta} \frac{\partial^2 V}{\partial x^2} \,.$$

The constants in the above equation are linked together by the Einstein relation, $D = \frac{T}{\eta}$, where T is the temperature. As usual, the equilibrium probability distribution reads $P_{eq}(x) \propto e^{-\frac{V(x)}{T}}$, and is the ground state eigenvector of the FP operator, $\bar{L}_{FP} P_{eq} = 0$.

In order to study the spectrum of $\bar{L}_{FP}$, it is useful to consider its Hermitian counterpart (see discussion in Sec. 3.1, and App. A.2), obtained as usual as:

$$\bar{H}_{FP} = \frac{1}{\phi_0(x)} \bar{L}_{FP}(x) \phi_0(x) = -D \frac{\partial^2}{\partial x^2} + V_{FP}(x,T) \,,$$

where $\phi_0(x) = \sqrt{P_{eq}(x)}$ and

$$V_{FP}(x,T) = \frac{D}{4T^2} \left(\frac{\partial V}{\partial x} \right)^2 - \frac{D}{2T} \frac{\partial^2 V}{\partial x^2} \,,$$

is an explicitly temperature dependent potential. The Hermitian operator $\bar{H}_{FP}$ is a quantum Hamiltonian which governs the spectrum and the dynamics of the FP problem. To appreciate the usefulness of such a transformation, consider the case in which the original potential is harmonic: $V(x) = \frac{1}{2} k x^2$. The transformed potential V_{FP} then reads:

$$V_{FP}(x,T) = D \left(\frac{k x}{2T} \right)^2 - \frac{D k}{2T} = \frac{1}{2} k' x^2 - \delta \,,$$

where $k' = \frac{D\,k^2}{2\,T^2}$ and $\delta = \frac{D\,k}{2\,T}$. The transformed potential is still harmonic, and its eigenvalues can be immediately written down as:

$$\lambda_n = \left(n + \frac{1}{2}\right)\,\omega - \delta = 2\,n\,\delta \qquad n = 0, 1, 2, \ldots$$

where $\omega = \sqrt{2\,D\,k'} = 2\,\delta$ is the harmonic frequency. Notice that $\lambda_0 = 0$. Since the levels are equispaced, the gap value is

$$\lambda_1(T) = 2\,\delta = \frac{D\,k}{T}\,. \tag{B.2}$$

B.2 Reduction to a two-level system

In general, a master equation can not be reduced to a simple Fokker-Planck form, as in Eq. B.1. In particular, the transition operator $\tilde{W}$ may be equal to $-\bar{L}_{FP}$ (see App. B.1) plus a perturbation (see also App. B.4). As a consequence, the hermitian counterpart of $\tilde{W}$, the Hamiltonian operator $\bar{H}$ (see Sec. 3.1), is also equal to $\bar{H}_{FP}$ (see App. B.1) plus a perturbation. In this case, by making use of the supersymmetric extension of the Fokker-Planck operator [75, 76], it is possible to prove that one can obtained a reduced operator which account only for low-lying spectrum of $\bar{H}$, provided that the perturbation does not change the ground state of $\bar{H}_{FP}$. The matrix elements of such a reduced operator can be computed considering only the hopping process between local minima. This great simplification clearly applies only at very small temperatures, when the equilibrium distribution $P_{eq}(x)$ is concentrated around the local minima ($x = x_\pm$ for the double-well potentials introduced in Sec. 2.2). In the case of the the double-well potential the reduced operator describes an effective *two level system* (TLS). Its correspondent (reduced) evolution operator reads:

$$\mathrm{P}\left[e^{-\Delta\,t\,\bar{H}}\right] = \begin{pmatrix} \langle x_+|e^{-\Delta\,t\,\bar{H}}|x_+\rangle & \langle x_+|e^{-\Delta\,t\,\bar{H}}|x_-\rangle \\ \langle x_-|e^{-\Delta\,t\,\bar{H}}|x_+\rangle & \langle x_-|e^{-\Delta\,t\,\bar{H}}|x_-\rangle \end{pmatrix}\,, \tag{B.3}$$

where P denotes the projector on the two-lowest states. This approach is equivalent to a semi-classical approximation which neglects Gaussian fluctuations around the classical solutions $x_{cl} = x_\pm$. [1]

Using the inverse of the transformation defined in Eq. 3.8, one can write down the correspondent reduced operator of $W = e^{-\Delta\,t\,\tilde{W}}$ (see Eq. 3.9):

$$\mathrm{P}\left[e^{t\,\tilde{W}}\right] = \begin{pmatrix} \langle x_+|e^{-\Delta\,t\,\bar{H}}|x_+\rangle & \langle x_+|e^{-\Delta\,t\,\bar{H}}|x_-\rangle\,e^{-\frac{V_+ - V_-}{2\,T}} \\ \langle x_-|e^{-\Delta\,t\,\bar{H}}|x_+\rangle\,e^{-\frac{V_- - V_+}{2\,T}} & \langle x_-|e^{-\Delta\,t\,\bar{H}}|x_-\rangle \end{pmatrix}\,. \tag{B.4}$$

[1] Gaussian fluctuations account for the harmonic modes in the bottom part of each well. Indeed that equilibrium distribution is locally a Gaussian near $x_\pm$, which is the well-known ground state solution of a Schrödinger equation in an harmonic potential. Therefore there are harmonic levels in the $\bar{H}$ spectrum and the first one accounts for the fastest intra-well dynamics (or relaxation). Nevertheless, at very low temperature, it can be neglected (or better: integrated-out) since the inter-well dynamics is the slowest one (see also App. B.1). This reasoning is true till an (effective) barrier between the wells is present (see Sec. 3.7).

Because the reduced evolution operator of the Master equation, defined in Eq. B.4, has to be stochastic (see Sec. 3.2) the following two conditions must hold:

$$\begin{aligned} \langle x_+|e^{-\Delta t\,\bar{H}}|x_+\rangle + \langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle\, e^{-\frac{V_- - V_+}{2\,T}} &= 1 \\ \langle x_+|e^{-\Delta t\,\bar{H}}|x_-\rangle\, e^{-\frac{V_+ - V_-}{2\,T}} + \langle x_-|e^{-\Delta t\,\bar{H}}|x_-\rangle &= 1\,. \end{aligned}$$

A further condition is that $\langle x_+|e^{-\Delta t\,\bar{H}}|x_-\rangle = \langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle$, since the matrix in Eq. B.3 is Hermitian. In conclusion, all the four entries of the reduced evolution operator, Eq. B.4, can be expressed in terms of only one independent quantity, the transition amplitude $\langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle$. The eigenvalues of reduced evolution operator, Eq. B.4, are simply given by:

$$\begin{aligned} \bar{\lambda}_0 &= 1 \\ \bar{\lambda}_1 &= 1 - 2\,\langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle\,\cosh(\frac{V_+ - V_-}{2\,T})\,, \end{aligned}$$

so that the its spectral gap reads: $\Delta = \bar{\lambda}_0 - \bar{\lambda}_1 = \langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle\, e^{-\frac{V_- - V_+}{2\,T}}\left(1 + e^{-\frac{V_+ - V_-}{T}}\right)$. Since T is assumed to be small, and $V_+ > V_-$, the second term in parenthesis is negligibly small. Therefore, the gap of the operator $\bar{H}$ is given by:

$$\Delta \simeq \langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle\, e^{-\frac{V_- - V_+}{2\,T}}\,. \tag{B.5}$$

The remaining part of this appendix is devoted to the calculation of the relevant transition amplitude $\langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle$. Once again, we start with a simple calculation, i.e., the Fokker-Planck case, where $\bar{H} = \bar{H}_{FP}$.

B.3 Calculation of transition amplitude: Fokker-Planck case

Here we want to compute the matrix element

$$\langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle\,,$$

in the Fokker-Planck case, for which the Hamiltonian operator is (see Sec. B.1):

$$\bar{H} = \bar{H}_{FP} = -D\,\nabla^2 + \frac{D}{4T^2}\left(\frac{\partial V}{\partial x}\right)^2 - \frac{D}{2\,T}\frac{\partial^2 V}{\partial x^2}.$$

It is simple enough to work out a Feynman Path-Integral version of this matrix element, which reads:

$$\langle x_-|e^{-\Delta t\,H_{loc}}|x_+\rangle = \int \mathcal{D}[x]\, e^{-\int_0^t \mathrm{d}t\,\mathcal{L}(x,\dot{x})}\,,$$

with the usual boundary conditions in imaginary-time, $x(0) = x_+$, $x(t) = x_-$. The Lagrangian is given by:

$$\mathcal{L}(x,\dot{x}) = \frac{1}{4\,D}\,\dot{x}^2 + \frac{D}{4T^2}\left(\frac{\partial V}{\partial x}\right)^2 - \frac{D}{2\,T}\frac{\partial^2 V}{\partial x^2}\,.$$

When the temperature is small, the last term in the above equation can be neglected, and we shall make use of the following approximation: $\mathcal{L}(x,\dot{x}) \simeq \frac{1}{4\,D}\,\dot{x}^2 + \frac{D}{4\,T^2}\left(\frac{\partial\,V}{\partial\,x}\right)^2$.

In order to obtain a semi-classical expansion [77] of this Feynman integral, one has to solve, as usual, the classical Lagrangian equations of motion:

$$\ddot{x} = \frac{D^2}{T^2}\,\frac{\partial\,V}{\partial\,x}\,\frac{\partial^2\,V}{\partial\,x^2},$$

A solution of this equation is given by:

$$\dot{x}_{cl} = \begin{cases} -\frac{D}{T}\,\frac{\partial\,V}{\partial\,x} & x \in [x_-, 0] \\ +\frac{D}{T}\,\frac{\partial\,V}{\partial\,x} & x \in [0, x_+] \end{cases} \quad . \tag{B.6}$$

This is the velocity of an overdamped motion [2] going from x_+ to x_-. Moreover in the space interval $[0, x_+]$ the system "climbs" the barrier, and so the force field is opposite to the velocity. After the top has been reached (with zero velocity), the system falls down onto the true minimum, and so in the interval $[x_-, 0]$ the force and the velocity have the same sign. We also note that the time-reversed trajectory $\bar{x}_{cl}(t) = x_{cl}(-t)$ is a classical solution of the Lagrangian equations of motion.

We can easily calculate the value of the action along the classical path, Eq. B.6. It is equal to:

$$\begin{aligned} S_{cl}[x_{cl}] &= \int_0^t \mathrm{d}t\,\mathcal{L}(x,\dot{x}) = \int_0^t \mathrm{d}t\left\{\frac{1}{4\,D}\,\dot{x}_{cl}^2 + \frac{D}{4\,T^2}\left(\frac{\partial\,V}{\partial\,x}\right)^2\right\} \\ &= \int_0^{\tilde{t}} \mathrm{d}t\left\{\frac{1}{4\,D}\,\dot{x}_{cl}^2 + \frac{D}{4\,T^2}\left(\frac{\partial\,V}{\partial\,x}\right)^2\right\} + \int_{\tilde{t}}^t \mathrm{d}t\left\{\frac{1}{4\,D}\,\dot{x}_{cl}^2 + \frac{D}{4\,T^2}\left(\frac{\partial\,V}{\partial\,x}\right)^2\right\}, \end{aligned}$$

where $\tilde{t}$ is defined by solving the equation $x_{cl}(\tilde{t}) = 0$. After some simplifications, we finally obtain:

$$\begin{aligned} S_{cl}[x_{cl}] &= \int_0^{\tilde{t}} \mathrm{d}t\,\frac{1}{2\,D}\dot{x}_{cl}\,\frac{\partial\,V}{\partial\,x} - \int_{\tilde{t}}^t \mathrm{d}t\,\frac{1}{2\,D}\dot{x}_{cl}\,\frac{\partial\,V}{\partial\,x} \\ &= \frac{1}{2\,T}\left(\int_{V_-}^{V_0} \mathrm{d}V - \int_{V_0}^{V_+} \mathrm{d}V\right) = \frac{1}{T}\left(V_0 - \frac{V_+ + V_-}{2}\right) = \frac{\bar{V}}{T}\,, \end{aligned}$$

where we used the notation $V_- = V(x_-)$ and $V_+ = V(x_+)$, and $\bar{V} = V_0 - \frac{V_+ + V_-}{2}$.

In conclusion, the semi-classical approximation leads to the following expression for the Fokker-Planck transition amplitude:

$$\langle x_-|e^{-\Delta\,t\,\bar{H}}|x_+\rangle \simeq C(t)\,e^{-\frac{\bar{V}}{T}}\,, \tag{B.7}$$

where $C(t)$ is a function that accounts for the Gaussian fluctuations around the classical solution. [78][page 1109ff]. It is not worth to know the analytical expression of such a coefficient, but it suffices to point out that it depends linearly on $\Delta\,t$. Indeed – since $x_+ \neq x_-$ – for small $\Delta\,t$ the transition matrix must be proportional to $\Delta\,t$:

$$\langle x_-|e^{-\Delta\,t\,\bar{H}}|x_+\rangle \simeq -\Delta\,t\,\langle x_-|\bar{H}|x_+\rangle\,.$$

[2] The velocity is proportional to the force.

Observe, in passing, that since the time-reversed path $\bar{x}_{cl}(t)$ from x_- to x_+ is also a classical solution, we can explicitly check that the transposed transition amplitude of the Hamiltonian reads:

$$\langle x_+|e^{-\Delta t\,\bar{H}}|x_-\rangle \simeq C(t)\,e^{-\frac{\bar{V}}{T}}\ .$$

The classical solution in Eq. B.6, and its time-reversed, are not the only possible solutions of the classical Lagrangian equations of motion. One can construct a whole family of paths bouncing many time between x_+ and x_-. For instance, consider the following path:

$$x_{cl}^{(3)} = \begin{cases} x_{cl}(t) & t \in [0, t_1] \\ \bar{x}_{cl}(t) & t \in [t_1, t_2] \\ x_{cl}(t) & t \in [t_2, t] \end{cases} ,$$

where by means of the expression $x_{cl}(t), t \in [0, t_1]$, we mean a classical solution, as defined in Eq. B.6, but with boundary condition $x_{cl}(0) = x_+$ and $x_{cl}(t_1) = x_-$. In a similar way $\bar{x}_{cl}(t)$ refers to the *time reversal* trajectory from $\bar{x}_{cl}(t_1) = x_-$ to $\bar{x}_{cl}(t_2) = x_+$. The last track is equivalent to the first, but with different boundary conditions. Because of the integration linearity, we easily find that:

$$S_{cl}[x_{cl}^{(3)}] = 3\,\frac{\bar{V}}{T}\ .$$

Summarizing, we conclude that (see also [78][page 180ff]):

$$\langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle \simeq \left(C(t)\,e^{-\frac{\bar{V}}{T}} + C_3(t)\,e^{-3\,\frac{\bar{V}}{T}} + \cdots\right) ,$$

where the series obviously goes on with higher-order terms obtained considering trajectories which a larger amount of back-and-forth trips. Nevertheless, the leading contribution comes uniquely from the first term, which is the one we shall retain. In the following, we shall refer to such an approximation as the *direct-path approximation*.

B.4 Calculation of transition amplitude: General case

For a pure Fokker-Planck equation the scale of the thermal fluctuation is $\sigma_{eff} \propto \sqrt{\frac{T}{k}}$ – where k is the (average) second derivative of the potential around the minima – and it is the only length-scale of the system. However, the proposal moves introduced in Sec. 3.3, bring in another length-scale: the proposal move range σ. Therefore two regimes are possible: Either $\sigma < \sigma_{eff}$ or $\sigma \geq \sigma_{eff}$.

In the first case, σ is the smallest scale and – if the temperature is small – this will justify the the derivation of an *effective Fokker-Planck*, just as in Sec. 3.4. If, on the other hand, σ_{eff} is the smallest scale (again for T small enough) we can split the full MC transition operator into two parts: An effective Fokker-Planck-like dynamics for displacements smaller than σ_{eff} plus a remaining *non-local* dynamics. As a consequence, the Hamiltonian operator $\bar{H}$ also splits into two terms, $\bar{H} = \bar{H}_{FP} + \bar{H}_{nloc}$. The first term in $\bar{H}$ is the form we have just considered in Sec. B.3, while the second term is a *non-local* part, where non-local means that $\langle x_1|H_{nloc}|x_2\rangle = 0$ if $|x_1 - x_2| < \sigma_{eff}$.

Using the Trotter's decomposition, the full transition amplitude can be written down in the following way:

$$\langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle = \int \prod_{i=1}^{N-1} \mathrm{d}x_i\, \langle x_N|e^{-\tau\,\bar{H}}|x_{N-1}\rangle \cdots \langle x_1|e^{-\tau\,\bar{H}}|x_0\rangle \ ,$$

where $\tau = \frac{\Delta t}{N}$ is the Trotter time-step, N is the number of Trotter slices, $x_0 = x_+$, and $x_N = x_-$. Moreover, if τ is small enough, one can substitute $\langle x_1|e^{-\tau\,\bar{H}}|x_0\rangle$ with $\langle x_1|e^{-\tau\,\bar{H}_{FP}}\,e^{-\tau\,\bar{H}_{nloc}}|x_0\rangle$, with an error $0(\tau^2)$. It is then possible to write the transition amplitude, by making an explicit use of this break-up, as:

$$\begin{aligned}\langle x_-|e^{-\Delta t\,\bar{H}}|x_+\rangle \quad &= \quad \int \prod_{i=1}^{2N-1} \mathrm{d}\,x_i\, \langle x_{2N}|e^{-\tau\,\bar{H}_{FP}}|x_{2N-1}\rangle\langle x_{2N-1}|e^{-\tau\,\bar{H}_{nloc}}|x_{2N-2}\rangle\cdots \\ &\qquad \langle x_2|e^{-\tau\,\bar{H}_{FP}}|x_1\rangle\langle x_1|e^{-\tau\,\bar{H}_{nloc}}|x_0\rangle \ .\end{aligned}$$

In the framework of the direct-path approximation introduced in Sec. B.3, only a single non-local jump is available during a path. Indeed, at very low temperatures the non-local move constraint $V(x_1) \simeq V(x_0)$ (see below) can be satisfied only by passing through the barrier, and this can be done just once. Using this approximation, we shall write down the transition amplitude in a more convenient way:

$$\begin{aligned}\langle x_-|e^{-\Delta t\,H}|x_+\rangle \quad &= \quad \int \frac{\mathrm{d}\,t_1\,\mathrm{d}t_2}{\tau}\,\delta(\Delta t - t_2 - t_1 - \tau) \int \mathrm{d}\,x_1\,\mathrm{d}\,x_2\,\langle x_-|e^{-\Delta t_2\,\bar{H}_{FP}}|x_2\rangle \\ &\qquad \langle x_2|e^{-\tau\,\bar{H}_{nloc}}|x_1\rangle\,\langle x_1|e^{-\Delta t_1\,\bar{H}_{FP}}|x_+\rangle \ ,\end{aligned}$$

where t_1, t_2 are the time elapsed by the local parts of the move and τ the same for the non-local one. The integrated coordinates x_1 and x_2 stand for the degrees of freedom of the starting and ending point of the single non-local jump. Inserting Eq. B.7 in the above expression, we obtain that:

$$\begin{aligned}\langle x_-|e^{-\Delta t\,H}|x_+\rangle \quad &= \quad \int \frac{\mathrm{d}\,t_1\,\mathrm{d}t_2}{\tau}\,\delta(\Delta t - t_2 - t_1 - \tau) \int \mathrm{d}x_1\,\mathrm{d}x_2\,C(t_2,x_2)\,e^{-\frac{1}{2T}\,(V(x_2)-V_-)} \\ &\qquad \langle x_2|e^{-\tau\,H_{nloc}}|x_1\rangle\,C(t_1,x_1)\,e^{-\frac{1}{2T}\,(V(x_1)-V_+)} \ .\end{aligned}$$

We emphasize that the Gaussian fluctuations around the classical solution depend also on the coordinates, because the latter are not fixed.

It remains a sort of ambiguity in the definition of τ, since the only "physical" time is $\Delta\,t$. However we observe that the above integral can be maximizes – and then such a problem solved – putting $\tau \to \Delta\,t$. Using this "variational" principle and the fact that $x_- \neq x_+$, one can approximate the *non-local transition amplitude* as follow:

$$\langle x_2|e^{-\Delta\,t\,\bar{H}_{nloc}}|x_1\rangle = T(x_2,x_1)\,e^{-\frac{1}{2T}\,|V(x_2)-V(x_1)|} \ ,$$

where $T(x_2,x_1)$ is the transition operator of a proposal move (e.g. Box, Gaussian or Lorentzian), that we defined in Sec. 3.3. [3] Changing the spatial integration variables, we obtain that:

$$\langle x_-|e^{-\Delta\,t\,\bar{H}}|x_+\rangle \quad = \quad \int \mathrm{d}\,x_1\,\mathrm{d}\,(x_2 - x_1)\,K_1(x_1,x_2;\Delta\,t)\,e^{-\frac{1}{2T}\,(V(x_2)+V(x_1)-V_+-V_-)}$$

[3] we have employed the symmetrization transformation for such an operator, as explained previously, and in Sec. 3.1.

$$T(|x_2 - x_1|)\, e^{-\frac{1}{2T}|V(x_2)-V(x_1)|} ,$$

where $K_1(x_1, x_2; \Delta t) = \int \frac{\mathrm{d}t_1\, \mathrm{d}t_2}{\Delta t}\, \delta(t_2 + t_1)\, C(x_1, t_1)\, C(x_2, t_2)$. (We made explicit use of the translation invariance of the proposal moves, $T(|x_2 - x_1|) = T(x_2, x_1)$, and we used, as usual, the notation $V_\pm = V(x_\pm)$.

At small temperature the exponential $e^{-\frac{1}{2T}|V(x_2)-V(x_1)|}$ is vanishingly small only if $V(x_2) \sim V(x_1)$. For every value of the distance $|x_2 - x_1|$, there is a unique value of the potential energy V such that $|x_2 - x_1| = d(V - V_+)$ (see Sec. 3.9.1 for a definition of the function d). It is still possible to use the saddle-point approximation to get rid of the integration in x_1, finding that:

$$\begin{aligned}\langle x_-|e^{-\Delta t\, \bar{H}}|x_+\rangle \quad &= \quad K_2 \int \mathrm{d}(x_2 - x_1)\, K_1(x_2 - x_1; \Delta t)\, e^{-\frac{1}{T}\left(d^{-1}(|x_2-x_1|)+\frac{V_+ + V_-}{2}\right)} \\ &\quad\; T(|x_2 - x_1|) ,\end{aligned}$$

where K_2 accounts for the Gaussian fluctuation around this saddle-point solution. [4] Performing a change of variables, making use the function d, we finally obtain that:

$$\langle x_-|e^{-\Delta t\, \bar{H}}|x_+\rangle = K_2 \int_{V_+}^{V_0} \mathrm{d}\, V\, \frac{\partial d}{\partial V}\, K_1(d(V - V_+); \Delta t)\, e^{-\frac{1}{T}\left(V - \frac{V_+ + V_-}{2}\right)}\, T(d(V - V_+)) \ . \quad \text{(B.8)}$$

The results obtained for the different trial moves are discussed in Sec. 3.9.2.

B.5 The harmonic excitations

In Sec. B.4 we defined the quantity $\sigma_{eff} \propto \sqrt{\frac{T}{k}}$, which is the range of the thermal fluctuations around a local minimum of a well, whose curvature (second derivative) is k. The other length-scale is the proposal move range σ. In this section we shall employ these length scales in order to discriminate between different regimes of the Monte Carlo Dynamics. We also recall that we introduced in Sec. 3.4 the jump moments of order j. We now extend the previous definition in a way that includes all the cases under inspection:

$$a_j(x) = \int \mathrm{d}y\, (y - x)^j\, \tilde{W}_{FP}(y, x),$$

where $\tilde{W}_{FP}(y, x) = \tilde{W}(y, x)$ if $\sigma < \sigma_{eff}$ while

$$\tilde{W}_{FP}(y, x) = \begin{cases} \tilde{W}(y, x) & |y - x| < \sigma_{eff} \\ 0 & \text{otherwise} \end{cases} ,$$

if $\sigma \geq \sigma_{eff}$. The first two jump moments enter in the definition of the Fokker-Planck equation (see Eq. 3.18). They are given by (see App. B.7):

$$a_1(x) = \begin{cases} -\frac{T_0}{3}\frac{\sigma^3}{T}\frac{\partial V}{\partial y} & \sigma < \sigma_{eff} \\ -\frac{T_0}{3}\frac{\sigma_{eff}^3}{T}\frac{\partial V}{\partial y} & \sigma \geq \sigma_{eff} \end{cases} ,$$

[4] We wrote $K_1(x_2 - x_1; \Delta t)$ instead of $K_1(x_2, x_1, \Delta t)$ because the Gaussian fluctuations are translational invariant.

and

$$a_2(x) = \begin{cases} \frac{2\,T_0}{3}\,\sigma^3 & \sigma < \sigma_{eff} \\ \frac{2\,T_0}{3}\,\sigma_{eff}^3 & \sigma \geq \sigma_{eff} \end{cases} ,$$

where $T_0 = \lim_{|y-x|\to 0} T(y,x)$. [5] Finally comparing Eq. 3.18 and Eq. B.1 we obtain that:

$$\begin{aligned} D &= \frac{1}{2}\,a_2(x), \\ \frac{1}{\eta} &= -\frac{a_1(x)}{\frac{\partial V}{\partial y}} , \end{aligned} \tag{B.9}$$

and we also note that the Einstein relation $D = \frac{T}{\eta}$ still holds.

In Sec. B.1 we found that the first non-vanishing eigenvalue of a Fokker-Planck equation with harmonic potential $V(x) = \frac{1}{2}\,k\,x^2$ is given by Eq. B.2. By substituting it in the previous value of the diffusion constant D we obtained that:

$$\lambda_1^{(harm)}(\sigma, T) \propto \begin{cases} \sigma^2\,\left(\frac{T}{k}\right)^{-1} & \sigma < \sigma_{eff} \\ \frac{1}{\sigma}\,\left(\frac{T}{k}\right)^{1/2} & \sigma \geq \sigma_{eff} \end{cases} . \tag{B.10}$$

From Refs. [75, 76] we know that, for a generic double-well potential, there are in the spectrum of the transition operator $\bar{H}$ (see Sec. 3.1) two low-lying states – the ground state and the fist exited state connecting the two minima – and a whole series of almost harmonic states which are due to the relaxation dynamics *inside* the two quasi-harmonic wells. Therefore the third eigenvalue λ_2 of a double-well potential should be equal to the first exited state of the harmonic potential, Eq. B.10.

We claimed in the Sec. 3.9 that there is an avoided crossing between λ_1 and λ_2 at very small temperature T or for a large value of the move range σ. Since both cases are compatible with the condition $\sigma \geq \sigma_{eff}$, we shall try to fit our previous data with this model.

In particular, we fit $\frac{\lambda_1}{T_0}$ versus σ from the data shown in Fig. 3.3, for large values of σ. From the theory, we know that $\frac{\lambda_1}{T_0} \propto T^{1/2}$. As one can see in Fig. B.1, there is an astonishing agreement between data and theory for the proposal moves considered so far. Moreover, one can also estimate that $\sigma_{eff} = 1.9\,\left(\frac{T}{k}\right)^{1/2}$, where the coefficient has been fitted from the numerical data.

We want to stress that – in contrast with the case of the λ_1 coefficient (see Eq. B.8) – the expression for the first harmonic gap is exactly known from the theory.

The avoided crossing between λ_1 and λ_2 occurs – at fixed temperature T – for a $\bar{\sigma}$ such that $\lambda_1(\bar{\sigma}, T) \simeq \lambda_2(\bar{\sigma}, T)$. In particular, for Box proposal move, this happens exactly when the effective barrier B_{eff} vanishes, at $\bar{\sigma} = \sigma_{cr} \simeq 1.75$, while for the other kind of moves (Gaussian, Lorentzian) $\bar{\sigma}$ depends on the temperature. This difference is due to the compact support of the Box move, which limits the long jumps to be within the move range.

B.6 The diffusion constant

Suppose a Markov chain was defined as like in Sec. 3.1. A configuration in the phase space is labeled by x and the average over the whole chain is indicated by angle brackets $\langle \dots \rangle$. We claim that the

[5] $T_0 = \frac{1}{2\,\sigma}, \frac{1}{\sqrt{2\,pi}\,\sigma}, \frac{1}{\pi\,\sigma}$ for Box, Gaussian, Lorentzian move respectively.

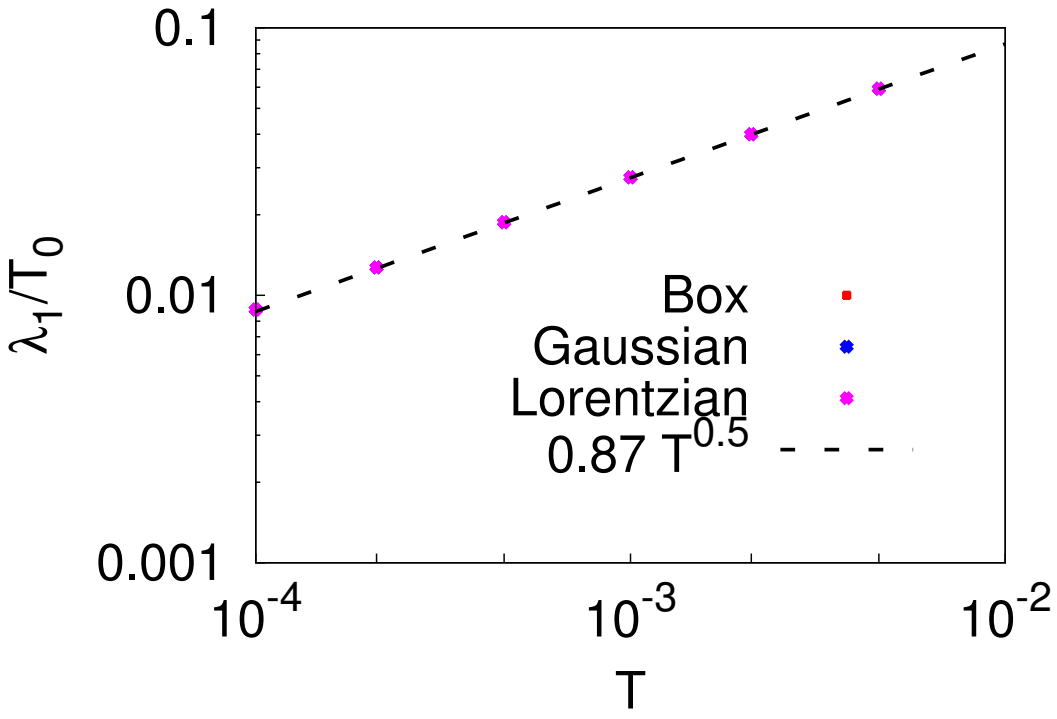

Figure B.1: Normalized gap values (see the text) versus the temperature, for Box Gaussian, and Lorentzian moves. These data fit with a $T^{1/2}$ dependence.

diffusion constant which appears in the Fokker-Plank equation, Eq. 2.2, is given by:

$$D = \lim_{\Delta t \to 0} \frac{\langle (x(t+\Delta t) - x(t))^2 \rangle}{2\,\Delta t} = \lim_{\Delta t \to 0} \frac{\langle x^2 \rangle - \langle x(t+\Delta t)\, x(t) \rangle}{\Delta t}\,, \tag{B.11}$$

and we shall prove that such a definition is consistent with the usual definition of D obtained by means of Eq. 3.18. We stress that – after the Markov chain reached the equilibrium – the process becomes stationary and so the diffusion constant does not longer depend on the time coordinate t, as in Eq B.11.

Using the symmetric transformation defined in Eq. 3.8, one can rewrite the time-dependent correlation function which appears in Eq B.11 as:

$$\langle x(t+\Delta t) x(t) \rangle = \langle \phi_0 | \hat{x}(t+\Delta t)\, \hat{x}(t) | \phi_0 \rangle\,,$$

where $\phi_0(x)$ is the ground state of the Hamiltonian operator associated to the Markov chain (see Sec. 3.1). We made use of hats in order to distinguish the position operator (in Heisenberg picture) from its eigenvalues (i.e. the positions x in real space).

By inserting two resolutions of the identity (in the energy basis) inside the previous expression, we find the following series expansion of the time-dependent correlation function:

$$\langle \phi_0 | \hat{x}(t)\, \hat{x}(0) | \phi_0 \rangle \quad = \quad \sum_{i,j} \langle \phi_0 | \hat{x} | \phi_i \rangle \langle \phi_i | e^{-\Delta t \lambda_i} | \phi_j \rangle \langle \phi_j | \hat{x} | \phi_0 \rangle$$

$$= \quad |\langle\phi_0|\hat{x}|\phi_0\rangle|^2 + \sum_{i>0} |\langle\phi_0|\hat{x}|\phi_i\rangle|^2 e^{-\Delta t \lambda_i} \ , \tag{B.12}$$

where $0 = \lambda_0 < \lambda_1 < \dots$ are the eigenvalues of the Hamiltonian defined in Eq. 3.8, which correspond to the eigenvectors $\{\phi_0, \phi_1, \dots\}$.

After a little algebra, the diffusion constant could be rewritten as:

$$D = \lim_{\Delta t \to 0} \frac{\langle x^2\rangle - \langle x\rangle^2 - \sum_{i>0} |\langle\phi_0|\hat{x}|\phi_i\rangle|^2 \, e^{-\Delta t \lambda_i}}{\Delta t} \ .$$

It is also simple to prove that the variance of the configuration along the Markov chain reads:

$$\langle x^2\rangle - \langle x\rangle^2 = \sum_{i>0} |\langle\phi_0|\hat{x}|\phi_i\rangle|^2 \ .$$

Finally we obtain a simple expansion of the diffusion constant as:

$$D = \lim_{\Delta t \to 0} \frac{\sum_{i>0} |\langle\phi_0|\hat{x}|\phi_i\rangle|^2 \left[1 - e^{-\Delta t \lambda_i}\right]}{\Delta t} = \sum_{i>0} \lambda_i |\langle\phi_0|\hat{x}|\phi_i\rangle|^2 \ . \tag{B.13}$$

A short comment is in order. In the case of an harmonic oscillator, the well known dipole selection rules yield to an even simpler result:

$$D_{harm} = \lambda_1 |\langle\phi_0|\hat{x}|\phi_1\rangle|^2 \ , \tag{B.14}$$

which implies that – in this case – the diffusion constant is proportional to the spectral gap λ_1.

For more general potentials this conclusion is no longer true, but one can hope that – in every practical situation – the diffusion constant D, as defined in Eq. B.11, is a good estimator of the relaxation properties of the system. Indeed, it is common wisdom to optimize the parameter of a Monte Carlo proposal move in such a way that the diffusion constant results to be maximized (see Sec. 3.5).

We report another analytical definition of the diffusion constant, equivalent to previous one. It reads:

$$D = -\frac{\partial}{\partial \Delta t} \langle x(t + \Delta t) \, x(t)\rangle = -\langle\phi_0|\hat{x} \, H \, \hat{x}|\phi_0\rangle \ . \tag{B.15}$$

This equation was obtained also employing Eqs. B.11 and B.12. H is again the Hamiltonian obtained by symmetrization of the transition operator defining the Markov chain. By inserting two resolutions of the identity (but in position space), we find the following expression of the diffusion constant:

$$\begin{aligned}
D \quad &= \quad -\int \mathrm{d}x \, \mathrm{d}y \, \langle\phi_0|\hat{x}|y\rangle\langle y|H|x\rangle\langle x|\hat{x}|\phi_0\rangle - \int \mathrm{d}x \, \mathrm{d}y \, \langle\phi_0|y\rangle y \langle y|H|x\rangle x\langle x|\phi_0\rangle \\
&= \quad \frac{1}{2}\int \mathrm{d}x \, \mathrm{d}y \, \langle\phi_0|y\rangle (y-x)^2 \langle y|H|x\rangle\langle x|\phi_0\rangle \\
&- \quad \frac{1}{2}\int \mathrm{d}x \, \mathrm{d}y \, \langle\phi_0|y\rangle (y^2 + x^2) \langle y|H|x\rangle\langle x|\phi_0\rangle \\
&= \quad \frac{1}{2}\int \mathrm{d}x \, \mathrm{d}y \, \langle\phi|y\rangle (y-x)^2 \langle y|H|x\rangle\langle x|\phi_0\rangle - \frac{1}{2}\int \mathrm{d}x \, x^2 \langle\phi_0|H|x\rangle\langle x|\phi_0\rangle + \\
&- \quad \frac{1}{2}\int \mathrm{d}y \, y^2 \langle\phi_0|y\rangle\langle y|H|\phi_0\rangle \ .
\end{aligned}$$

The last two terms are zero because H acts directly on ground state ϕ_0 . [6]

Therefore, the diffusion constant is given by:

$$D = \frac{1}{2}\int dx\, dy\, \langle\phi_0|y\rangle (y-x)^2 \langle y|H|x\rangle\langle x|\phi_0\rangle \tag{B.16}$$

$$= \frac{1}{2}\int dx\, dy\, (y-x)^2 W(y,x)\, P_{eq}(x)\ . \tag{B.17}$$

where $W(y,x) = \frac{\langle\phi_0|y\rangle}{\langle\phi_0|x\rangle}\langle y|H|x\rangle$ is the inverse of the symmetric transform introduced in Eq. 3.8 and $P_{eq}(x)$ is the equilibrium distribution of the Markov chain.

Using the notation of the Sec. 3.4, we finally attain that:

$$D = \frac{1}{2}\int \mathrm{d}x\, a_2(x)\, P_{eq}(x)\ ,$$

where $a_2(x)$ is the second jump moment (see Eq. 3.18). We emphasize that, whether the jump moment does not depend on the position or it varies over a range larger than the range of the thermal fluctuations, one can safely use the approximation: $D \simeq \frac{a_2}{2}$, as we wanted to prove.

B.7 Jump moments

In this section we shall sketch the derivation of the jump moments as they appear in App. B.5. We remember that the jump moment of order k is defined as:

$$a_k(x) = \int_{-\infty}^{+\infty} \mathrm{d}y\, (y-x)^k\, T(y,x) \min\left[1, e^{-\frac{V(y)-V(x)}{T}}\right]\ ,$$

where $T(y,x)$ is the transition operator related to the Monte Carlo proposal move (see Sec. 3.4 for a thorough introduction).

First of all, we first all suppose that the argument of the integral is non-vanishing only if $|y-x| < \sigma_{eff}$, and also that σ_{eff} is an absolute small quantity. These assumptions are reasonable in the context of the proposal moves considered in Sec. 3.5 if the temperature T is not too much large. Therefore we can rewrite the above equation as:

$$a_k(x) \simeq \int_{x-\sigma_{eff}}^{x+\sigma_{eff}} \mathrm{d}y\, (y-x)^k\, T(y,x) \min\left[1, e^{-\frac{V(y)-V(x)}{T}}\right]\ .$$

Moreover – since all the proposal move considered in Sec. 3.5 are symmetric – we can employ the further approximation: $T(y,x) = T_0 + C_2\,(y-x)^k + 0\left(\sigma_{eff}^4\right)$, and put it in the definition of the jump moments. It results that:

$$a_k(x) \simeq T_0\, b_k(x) + C_1\, b_{k+2}(x) + 0\left(\sigma_{eff}^4\right)\ ,$$

where $b_k(x) = \int_{x-\sigma_{eff}}^{x+\sigma_{eff}} \mathrm{d}y\, (y-x)^k \min\left[1, e^{-\frac{v(y)-V(x)}{T}}\right]$, and C_1 an unknown coefficient.

[6] Recall that the ground state energy λ_0 of the Hamiltonian is zero (see Sec.3.1).

Then, we take a last approximation. Provided that $F(x) = -\frac{\partial V}{\partial x} \neq 0$, one obtains that

$$b_k(x) \simeq \int_{x-\sigma_{eff}}^{x+\sigma_{eff}} \mathrm{d}y\,(y-x)^k \min\left[1, e^{+\frac{F(x)(y-x)}{T}}\right] .$$

The remaining singular points – where $F(x) = 0$ – will be finally included in our formulation since all the function that we employed are continuous.

Without loss of generality, we assume that $F(x) < 0$. As a consequence we can expand the previous equation as follow:

$$b_k(x) \simeq \int_{x-\sigma_{eff}}^{x+\sigma_{eff}} \mathrm{d}y\,(y-x)^k + \int_{x-\sigma_{eff}}^{x+\sigma_{eff}} \mathrm{d}y\,(y-x)^k\, e^{+\frac{F(x)(y-x)}{T}} .$$

and – by means of integration by part – we find the following recursive relation:

$$b_k(x) = (-)^k \frac{\sigma_{eff}^{k+1}}{k+1} + \frac{T}{F(x)} e^{\frac{F(x)\sigma_{eff}}{T}} + \frac{T}{F(x)}\, k \left(b^{k-1}(x) + (-)^k \frac{\sigma_{eff}^k}{k} \right) \qquad k > 0 .$$

It is very simple to work out the zero-order jump moment, which reads:

$$b_0(x) = \sigma_{eff} + \frac{T}{F(x)} \left(e^{\frac{F(x)\sigma_{eff}}{T}} - 1 \right) \simeq 2\sigma_{eff} + 0\left(\sigma_{eff}^2\right) ,$$

and, by means of the above recursive relation, all the others. We report only the first two because they are the sole which appear in the Fokker-Plank equation (see Eq.3.18). They read:

$$\begin{aligned} b_1(x) &= \frac{1}{3} \frac{F(x)}{T} \sigma_{eff}^3 + O\left(\sigma_{eff}^4\right) \\ b_2(x) &= \frac{2}{3} \sigma_{eff}^3 + O\left(\sigma_{eff}^4\right) . \end{aligned}$$

Finally, we obtain we following expression of the jump moments:

$$\begin{aligned} a_0(x) &= 2\,T_0 \sigma_{eff} + O\left(\sigma_{eff}^2\right) \\ a_1(x) &= \frac{T_0}{3} \frac{F(x)}{T} \sigma_{eff}^3 + O\left(\sigma_{eff}^4\right) \\ a_2(x) &= \frac{2\,T_0}{3} \sigma_{eff}^3 + O\left(\sigma_{eff}^4\right) . \end{aligned}$$

A few final comments are in order. The fact that $a_0(x) \propto \sigma_{eff} < 1$ is not in contradiction with the normalization of the transition operator. Indeed, in the definition of such an operator, one did not include the singular part (proportional to a Dirac's delta function) which account for move rejected during a Metropolis Monte Carlo step (see Sec. 3.2). More in general, all the jump moments go to zero as a power of σ_{eff} since the measure of the subset of the possible final configurations y (easy) attainable from an initial configuration x shrinks according to that parameter.

Appendix C

Appendix of Chapter 4

C.1 The Primitive Approximation for the action

The Eq. 4.2 can be directly rewritten as a product of P terms as follow:

$$Z(\beta) = \mathrm{Tr}\left[\underbrace{e^{-\frac{\beta}{P}H}\cdots e^{-\frac{\beta}{P}H}}_{P}\right]. \tag{C.1}$$

The number P of such terms is also called the number of *Trotter slices*. Every exponential in Eq. C.1 is similar to the original exponential of the Hamiltonian, but now with a reduced coefficient $\beta_{eff} = \beta/P$. Pushing this idea a step further, it is possible to define an *effective temperature* $T_{eff} = PT$, which is *larger* than the original one: the Feynman approach aims at taking advantage of such a *high temperature* approximation in order to estimate the partition function. Let us insert $P-1$ resolutions of the identity ($1 = \int \mathrm{d}x\,|x\rangle\langle x|$) between every term in Eq. C.1:

$$Z(\beta) = \int \prod_{i=1}^{P-1} \mathrm{d}x_i\ \mathrm{Tr}\left[\underbrace{e^{-\frac{\beta}{P}H}|x_{P-1}\rangle\langle x_{P-1}|\cdots|x_1\rangle\langle x_1|e^{-\frac{\beta}{P}H}}_{P}\right]. \tag{C.2}$$

It is worth to introduce the so-called *density operator*:

$$\rho\left(x', x; \frac{\Delta t}{\hbar}\right) = \langle x'|e^{-\frac{\Delta t}{\hbar}H}|x\rangle,$$

where $\frac{\Delta_T}{\hbar} = \beta/P$. For the sake of simplicity, from now on we shall set $\hbar = 1$. In doing so, we have moved from the old notation with the temperature $1/\beta$, to the new one with the time-interval Δt because we want to stress the equivalence between this Path-Integral and that of quantum-field theory in imaginary-time (i.e., after the so-called Wick rotation, see for instance [79]). In future development, we shall keep this last notation because it leads to equations which are easier to remember. By inserting the expression of the density operator inside Eq. C.2, we obtain a formal product that is written as a

convolution integral:

$$Z(\Delta t) = \int \prod_{i=0}^{P-1} \mathrm{d}x_i \, \rho\,(x_{i+1}, x_i; \Delta t)\,, \quad \text{where} \quad x_P = x_0\ . \tag{C.3}$$

Until now, everything is general. We know specify the Hamiltonian to be the usual quadratic one:

$$H = \frac{p^2}{2m} + V\ ,$$

where $p = -i\,\partial_x$. As in Chap. 2, $\Gamma = \frac{1}{2m}$ (remember that $\hbar = 1$) will be our annealing parameter. The kinetic term $\hat{T} = p^2/(2m)$ does not commute with the potential $\hat{V}$ and it provides the quantum fluctuations which cause a test particle to eventually tunnel through the barriers. With these notations, the density operator is:

$$\rho\,(x', x; \Delta t) = \langle x'|e^{-\Delta t\,(\hat{T}+\hat{V})}|x\rangle\ ,$$

and, observing that for small Δt the following equation holds

$$e^{-\Delta t\,(\hat{T}+\hat{V})} = e^{-\Delta t\,\frac{\hat{V}}{2}}\, e^{-\Delta t\,\hat{T}}\, e^{-\Delta t\,\frac{\hat{V}}{2}} + O\left(\Delta t^3\right) \tag{C.4}$$

we obtain that:

$$\rho\,(x', x; \Delta t) = \langle x'|e^{-\Delta t\,\hat{T}}|x\rangle\, e^{-\Delta t\,\frac{V(x')+V(x)}{2}} + O\left(\Delta t^3\right)\ . \tag{C.5}$$

Eq. C.4 is equivalent to a first-order approximation of the so-called *Trotter formula*

$$e^{-\beta\,(\hat{T}+\hat{V})} = \lim_{P\to\infty} \left(e^{-\frac{\beta}{P}\hat{T}}\, e^{-\frac{\beta}{P}\hat{V}}\right)^P\ ,$$

which is Hermitian, as the original exponential operator. This is a subtle but important point, because such an Hermitian approximation is correct up to *second order*, while the argument of the single exponentials in the R.H.S. of Eq. C.4 contains only first-order terms. This increase of the precision is a fingerprint of a whole class of Hermitian approximations, as we shall see in Sec. C.2.

It is now possible to work out analytically the matrix element in Eq. C.5 (e.g., by means of a Fourier transform) to obtain:

$$\rho\,(x', x; \Delta t) = \left(\frac{1}{4\,\pi\,\Gamma\,\Delta t}\right)^{\frac{1}{2}}\, e^{-\Delta t\left\{\frac{m}{2}\left(\frac{x'-x}{\Delta t}\right)^2 + \frac{V(x')+V(x)}{2}\right\}} + O\left(\Delta t^3\right)\ .$$

Inserting into Eq. C.3, we find that the partition function reads:

$$Z(\beta) = \left(\frac{1}{4\,\pi\,\Gamma\,\Delta t}\right)^{\frac{P}{2}} \int \prod_{i=0}^{P-1} \mathrm{d}x_i\, e^{-S_{PA}[x]} + O\left(\Delta t^3\, P\right)\ , \tag{C.6}$$

where $S_{PA}[x]$ is the so-called (euclidean) *primitive action*:

$$S_{PA}[x] = \sum_{i=0}^{P-1} \Delta t \left\{\frac{m}{2}\left(\frac{x_{i+1}-x_i}{\Delta t}\right)^2 + V(x_i)\right\} \quad \text{where} \quad x_P = x_0\ . \tag{C.7}$$

We want to make a few comments about this approximation. First of all, in the limit $\Delta t \to 0$ ($P \to \infty$) $\{x_i\}$ becomes a trajectory in the configuration space labeled by the "time" t. This trajectory is *closed* because $x(\beta) = x(0)$, as required by the Trace condition. Therefore $S_{PA}[x]$ in Eq. C.7 becomes the (euclidean) action $\int \mathrm{d}t\, H(x(t))$, with periodic boundary condition. In Fig. C.1 a configuration of the PIMC is sketched and it may be helpful for visualizing the inter-slice couplings scheme of Eq. C.7.

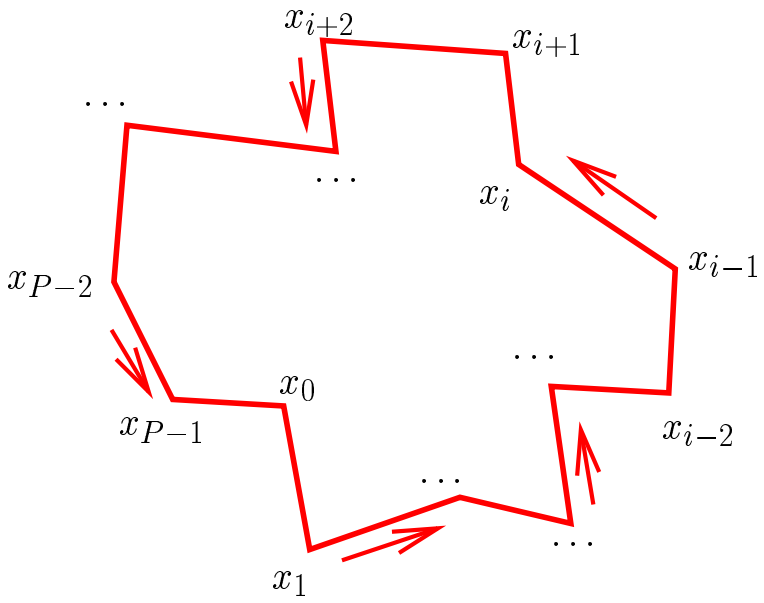

Figure C.1: Pictorial view of the PIMC configuration. One can see how the Trotter's slices are arranged in a closed chain. Bonds represent the kinetic part of Eq. C.7. By means of the arrows we indicate the orientation of the chain (according to the growing subscript of the Trotter's slices).

The quantum partition function written as in Eq. C.6 is indeed a classical partition function for P copies of the original system. All the copies feel the same (original) potential energy $V(x)$ and interact by means of an harmonic potential whose strength depends on the number of copies itself and on the annealing parameter Γ. Letting $\Gamma \to 0$ – during the quantum annealing simulation – the stiffness of such springs will diverge. Therefore, for very small values of Γ (at the very end of the annealing) all the copies are forced to have the same configuration x. This is nothing but the *classical limit* of the original quantum system.

As for the Trotter discretization, one can show that, for instance, the internal energy has an error which is $0\left(\Delta t^2\right) = 0\left(\beta/P\right)^2$. This means that, for any given temperature, by increasing the number of Trotter's slices we should observe a quadratic convergence to the exact value. Finally we notice that for a fixed target precision of a simulation, the number of Trotter's slices must be proportional to the inverse of the temperature. This is indeed a drawback of the PIMC: exploring regions of very small

temperature is very time expensive, being the computational time required by a PIMC algorithm (at least) linear in P.

C.2 Fourth-Order approximation

Not surprisingly, much effort has been devoted to searching better approximations of the exact action than the primitive one. A general theory – which includes also applications to symplectic integrators employed in molecular dynamics – has been formulated by Suzuki [80, 81, 82]. In general, this approach makes use of many nested commutators and is best suited to lattice problem [83]. A simpler formulation has been found by Takahashi and Imada [84]. Their *fourth-order* approximation involves only local operators and it is remarkably equal to the semi-classical expansion of the partition function [85, 86]. Recently, it has been pointed out that such an approximation does not include all the fourth-order terms of the actual expansion [87, 88]. Nevertheless, the Takahashi-Imada approximation (TIA) provided very good results in quantum fluid simulations [89].

The Takahashi-Imada approximation of the partition function reads:

$$\begin{aligned} \mathrm{Tr}\left[e^{-\beta(\hat{T}+\hat{V})}\right] &= \mathrm{Tr}\left[e^{-\frac{\beta}{P}\hat{T}}e^{-\frac{\beta}{P}\hat{V}_{eff}}\right]^P + O\left(\Delta t^5\, P\right) \\ \hat{V}_{eff} &= \hat{V} + \frac{1}{24}\left(\frac{\beta}{P}\right)^2 [\hat{V},[\hat{T},\hat{V}]]\,. \end{aligned} \tag{C.8}$$

(For a proof, see the original literature.) Because we shall apply it only to a one-dimensional potential, for the sake of simplicity we shall study only the equations for this case.

The effective potential V_{eff}, Eq. C.8, is a correction of the bare one which depends on the number of Trotter's slices. Since the double commutator is easily expanded in [85]:

$$[\hat{V},[\hat{T},\hat{V}]] = 2\,\Gamma\,\left(\frac{\partial\, V(x)}{\partial\, x}\right)^2\,,$$

we obtain that:

$$V_{eff}(x) = V(x) + \frac{1}{12}\Gamma\,(\Delta\, t)^2\,\left(\frac{\partial V(x)}{\partial\, x}\right)^2\,. \tag{C.9}$$

Obviously, for $\Delta\, t \to 0$ $V_{eff} \to V(x)$, so that both the primitive and the Takahashi-Imada schemes lead to the same continuous limit. Moreover, because the non-trivial part of $V_{eff}(x)$ is always positive and acts only when the variation of the potential is not negligible, it causes the particle to stay longer around the local minima of the bare potential. The price to pay is that the computation of the forth-order approximation is heavier than the primitive one due to the extra term added in Eq. C.9. In particular, the knowledge of the first derivative of the potential is required.

The fourth-order approximation to the partition function reads:

$$\begin{aligned} Z(\beta) &= \left(\frac{1}{4\,\pi\,\Gamma\,\Delta\, t}\right)^{\frac{P}{2}} \int \prod_{i=0}^{P-1} \mathrm{d}x_i\, e^{-S_{TIA}[x]} + 0\left(\Delta\, t^5\, P\right), \\ S_{TIA}[x] &= \sum_{i=0}^{P-1} \Delta\, t\left\{\frac{m}{2}\left(\frac{x_{i+1}-x_i}{\Delta\, t}\right)^2 + V_{eff}(x_i)\right\} \quad \text{where} \quad x_P = x_0\,. \end{aligned} \tag{C.10}$$

All the statistical averages for the observables can be carried out as usual. Obviously, the thermodynamical estimator of the internal energy has a smaller error $0\left(\Delta t^4\right) = 0\left(\frac{\beta}{P}\right)^4$.

C.3 Statistical estimators

Even though the Trotter error is good, big statistical errors could arise from a Monte Carlo sampling of poor observable estimators, and this can completely spoil the quality of the approximation. For instance, the thermodynamical internal energy estimators in the case of the primitive and fourth-order approximation read:

$$\begin{aligned} U_{PA} &= \frac{1}{2\,\Delta t} - \frac{1}{4\,\Gamma}\sum_{i=0}^{P-1}\frac{1}{P}\left(\frac{x_{i+1}-x_i}{\Delta t}\right)^2 + \frac{1}{P}\sum_{i=0}^{P-1} V_{eff}(x_i) \\ U_{TIA} &= U_{PA} + \frac{\Gamma}{4}\frac{\Delta t^2}{P}\sum_{i=0}^{P-1}\left(\frac{\partial V(x_i)}{\partial x_i}\right)^2 . \end{aligned}$$

The true internal energy is clearly:

$$\begin{aligned} U &= \langle U_{PA}\rangle + O\left(\Delta t\right)^2 \\ U &= \langle U_{TIA}\rangle + O\left(\Delta t\right)^4 . \end{aligned}$$

However, the Monte Carlo sampling of both these estimator leads to unreliable results. Indeed the first two terms of U_{PA} almost cancel out [59] and this cancellation gives rise to large fluctuations which usually increase with the number of Trotter's slices [90] because of their $1/\Delta t$ dependence. [1] We now check that these fluctuations enter only in the kinetic term. In order to find an expression for the potential energy estimator, consider a slightly different Hamiltonian $H = \hat{T} + \alpha\,\hat{V}$. The potential energy can then be expressed as:

$$\langle V\rangle = -\frac{1}{\beta}\frac{\partial}{\partial\,\alpha}\ln Z(\beta,\alpha)|_{\alpha=1} .$$

According to this formula, we obtain the following potential energy estimators for the two approximations:

$$\begin{aligned} V_{PA} &= \frac{1}{P}\sum_{i=1}^{P} V(x_i) \\ V_{TIA} &= V_{PA} + \frac{\Gamma}{6}\frac{\Delta t^2}{P}\sum_{i=0}^{P-1}\left(\frac{\partial V(x_i)}{\partial x_i}\right)^2 . \end{aligned} \tag{C.11}$$

Taking the difference of U and V, we obtain the following estimators for the kinetic energy:

$$K_{PA} = \frac{1}{2\,\Delta t} - \frac{1}{4\,\Gamma\,P}\sum_{i=0}^{P-1}\left(\frac{x_{i+1}-x_i}{\Delta t}\right)^2$$

[1] Recall that in a free random walk, i.e. when $V=0$, $\sum_{i=0}^{P-1}\left(x_{i+1}-x_i\right)^2 \propto P$.

$$K_{TIA} = K_{PA} + \frac{\Gamma}{12}\frac{\Delta t^2}{P}\sum_{i=0}^{P-1}\left(\frac{\partial V(x_i)}{\partial x_i}\right)^2 . \tag{C.12}$$

We notice that a contribution involving the potential gradient (i.e. the force) is also present in the TIA kinetic estimator.

Actually, the observable estimators are not completely determined by the actions only. One can consider the following equality:

$$\frac{\int \prod_{i=0}^{P-1} \mathrm{d}x_i \left\{\sum_{i=0}^{P-1} (x_i - \bar{x}) \frac{\partial S[x]}{\partial x_i}\right\} e^{-S[x]}}{\int \prod_{i=0}^{P-1} \mathrm{d}x_i \, e^{-S[x]}} = P - 1 \tag{C.13}$$

where $S[x]$ stands for either the Primitive or Takahashi-Imada actions, Eq. C.7 and Eq. C.10, while $\bar{x} = \frac{1}{P}\sum_{i=0}^{P-1} x_i$ is the *centroid* coordinate. Eq. C.13 was obtained by means of an integration by parts, and it is very similar to the virial in classical mechanics [91][page 22ff]. Substituting the action inside such an equation, it becomes

$$P - 1 = \begin{cases} \Delta t \left\{ \frac{1}{2\Gamma} \sum_{i=0}^{P-1} \left(\frac{x_{i+1}-x_i}{\Delta t}\right)^2 + \sum_{i=0}^{P-1} (x_i - \bar{x}) \frac{\partial V(x_i)}{\partial x_i} \right\} & \text{PA} \\ \Delta t \left\{ \frac{1}{2\Gamma} \sum_{i=0}^{P-1} \left(\frac{x_{i+1}-x_i}{\Delta t}\right)^2 + \sum_{i=0}^{P-1} (x_i - \bar{x}) \frac{\partial V_{eff}(x_i)}{\partial x_i} \right\} & \text{TIA.} \end{cases}$$

From Eq. C.14 one then obtain that:

$$\frac{1}{4\Gamma P}\sum_{i=0}^{P-1}\left(\frac{x_{i+1}-x_i}{\Delta t}\right)^2 = \begin{cases} \frac{P-1}{2P\Delta t} - \frac{1}{2}\sum_{i=0}^{P-1} (x_i - \bar{x}) \frac{\partial V(x_i)}{\partial x_i} & \text{PA} \\ \frac{P-1}{2P\Delta t} - \frac{1}{2}\sum_{i=0}^{P-1} (x_i - \bar{x}) \frac{\partial V_{eff}(x_i)}{\partial x_i} & \text{TIA} \end{cases}$$

Finally, putting Eq. C.14 into Eq. C.12, we obtain the so-called *virial estimators* of the kinetic energy:

$$\begin{aligned} \tilde{K}_{PA} &= \frac{1}{2\beta} + \frac{1}{2P}\sum_{i=0}^{P-1} (x_i - \bar{x}) \frac{\partial V(x_i)}{\partial x_i} \\ \tilde{K}_{TIA} &= \tilde{K}_{PA} + \frac{\Gamma}{12}\frac{\Delta t^2}{P}\sum_{i=0}^{P-1} (x_i - \bar{x}) \frac{\partial V(x_i)}{\partial x_i} \frac{\partial^2 V(x_i)}{\partial x_i^2} . \end{aligned} \tag{C.14}$$

We stress that no dangerous cancellations occur in Eqs. C.14, and their classical limit is $\tilde{K} = \frac{1}{2\beta}$, has it should be, due to equipartition. Quantum corrections to classical equipartition appear to depend on the fluctuations $(x_i - \bar{x})$, which are proportional to the *De Broglie's thermal length* $\sqrt{4\pi\Gamma\beta}$ (in the case of a smooth potential) [59]. The potential energy is not affected by the virial transformation, and Eqs. C.11 are still valid. There is a trade-off in the game: the virial estimator requires a further derivative of the potential. In the most accurate case – Takahashi-Imada approximation plus virial estimator – one has to compute up to the second-order derivative of the potential.

C.4 Sampling methods

In the previous sections, we studied several approximations of the partition function and several estimators of the main observables. A discussion of the sampling methods is now in order. The aim

of the Path-Integral Monte Carlo is sampling configurations taken over a suitable (finite dimensional) subspace of the whole Hilbert space. Since the partition function is written as:

$$Z = \left(\frac{1}{4\,\pi\,\Gamma\,\Delta\,t}\right)^{\frac{P}{2}} \int \prod_{i=0}^{P-1} \mathrm{d}x_i\, e^{-S[x]},$$

this is equivalent to sampling vectors of $\mathbf{R}^P$ distributed according to the multivariate distribution:

$$P_{eq}(x) = \left(\frac{1}{4\,\pi\,\Gamma\,\Delta\,t}\right)^{\frac{P}{2}} e^{-S[x]}\,. \tag{C.15}$$

Since we shall employ the Metropolis algorithm, we use the notation of Sec. 3.2. A trial move $T(x', x; \Delta\,t)$ may involve many Trotter slices (i.e. component of the vector x), or just a single one (viz. a *single slice move*). Anyway – instead of proposing a completely random displacement – one can try to sample the kinetic part of the Hamiltonian exactly. We recall that the kinetic part of the distribution defined in Eq. C.15 is given by:

$$P_K(x) = \left(\frac{1}{4\,\pi\,\Gamma\,\Delta\,t}\right)^{\frac{P}{2}} \prod_{i=0}^{P-1} e^{-\Delta\,t\left\{\frac{1}{4\,\Gamma}\left(\frac{x_{i+1}-x_i}{\Delta\,t}\right)^2\right\}},$$

and it is formally the trace of a Gaussian convolution. If we set:

$$T_K(x', x) = \frac{P_K(x')}{\int \prod_{i=0}^{P-1} \mathrm{d}x'_i\, P_K(x')},$$

we shall obtain a so-called *heat-bath algorithm* for a free (i.e. $V(x) = 0$) system. The acceptance factor reads, in this case, as:

$$A_K(x', x) = \min\left(1,\, \frac{T_K(x, x')\,P_K(x')}{T_K(x', x)\,P_K(x)}\right) = 1\,,$$

which is the best possible for a Metropolis algorithm. When we switch on the potential V this result will be no longer true:

$$A(x', x) = \min\left(1,\, \frac{T_K(x, x')\,P_{eq}(x')}{T_K(x', x)\,P_{eq}(x)}\right) = \begin{cases} \min\left(1, \frac{\prod_{i=0}^{P-1} e^{-\Delta\,t\,V(x'_i)}}{\prod_{i=0}^{P-1} e^{-\Delta\,t\,V(x_i)}}\right) & \text{PA} \\ \min\left(1, \frac{\prod_{i=0}^{P-1} e^{-\Delta\,t\,V_{eff}(x'_i)}}{\prod_{i=0}^{P-1} e^{-\Delta\,t\,V_{eff}(x_i)}}\right) & \text{TIA} \end{cases},$$

but one can reasonably expect that, for a smooth potential, the acceptance will remain reasonably high.

The simplest illustration of such a proposal move involves only a single-slice displacement. Unfortunately, this is seldom a good choice. Indeed, a single slice change scarcely affects the global configuration, and the correlation time is expected to be very long. Although it is not the better choice, it is worth to start from such an example before to approach more effective and global proposal moves. Indeed, changing only a single x_i, one can neglect all the others slices but the nearest ones, x_{i+1} and x_{i-1}, because they cancel out in the ratio $\frac{P_K(x')}{P_K(x)}$. Therefore, we obtain that:

$$T_K^{(1)}(x'_i, x) = \frac{\left(\frac{1}{4\,\pi\,\Gamma\,\Delta\,t}\right) e^{-\Delta\,t\left\{\frac{1}{4\,\Gamma}\left(\frac{x_{i+1}-x'_i}{\Delta\,t}\right)^2\right\}} e^{-\Delta\,t\left\{\frac{1}{4\,\Gamma}\left(\frac{x'_i-x_{i-1}}{\Delta\,t}\right)^2\right\}}}{\left(\frac{1}{4\,\pi\,\Gamma\,\Delta\,t}\right) \int \mathrm{d}x'_i\, e^{-\Delta\,t\left\{\frac{1}{4\,\Gamma}\left(\frac{x_{i+1}-x'_i}{\Delta\,t}\right)^2\right\}} e^{-\Delta\,t\left\{\frac{1}{4\,\Gamma}\left(\frac{x'_i-x_{i-1}}{\Delta\,t}\right)^2\right\}}}$$

$$= \quad (2\,\pi\,\Delta\,t\,\Gamma)^{-\frac{1}{2}}\; e^{-\frac{\left(x'_i - \frac{x_{i+1}+x_{i-1}}{2}\right)^2}{2\,\Delta\,t\,\Gamma}}\;,$$

which is a Gaussian of mean $\frac{x_{i+1}+x_{i-1}}{2}$ and variance $\Delta\,t\,\Gamma$. This is a really non-trivial property of the Gaussian distribution, and it is not shared by other distributions which are also stable under convolution, like the Lorentzian or more general Lévy distributions [92, 93]. Sampling a Gaussian is a very fast process [53, 74] on a computer, but many Monte Carlo steps must be performed in order to have decorrelated configurations. Usually, a single Path-Integral MC step consists, by definition, of a move for each particle and for each Trotter slice (on average). The order of the single slice moves can be either sequential or random, even though the second choice guarantees better decorrelation.

C.5 The Lévy construction

The simple single-slice move of the previous section is the basis of a more effective strategy which involves many slices at time. The goal is still to find an algorithm which samples exactly the kinetic part of $P_{eq}(x)$ in an iterative way. First of all, we need a few more notations. The kinetic part of the density operator Eq. C.5 is given by:

$$\rho_K\,(x', x; \Delta\,t) = \left(\frac{1}{4\,\pi\,\Gamma\,\Delta\,t}\right)^{\frac{1}{2}}\; e^{-\Delta\,t\left\{\frac{1}{4\,\Gamma}\left(\frac{x'-x}{\Delta\,t}\right)^2\right\}}, \tag{C.16}$$

and, as a consequence, the kinetic part of $P_{eq}(x)$ reads:

$$P_K(x) = \prod_{i=0}^{P-1}\,\rho_K\,(x_{i+1}, x_i; \Delta\,t)\;\;,$$

with the usual periodic boundary condition $x_P = x_0$. The single-slice transition probability can be written as:

$$T_K^{(1)}(x'_i, x) = \frac{\rho_K\,(x_{i+1}, x'_i; \Delta\,t)\;\rho_K\,(x'_i, x_{i-1}; \Delta\,t)}{\rho_K\,(x_{i+1}, x_{i-1}; 2\,\Delta\,t)}\;. \tag{C.17}$$

It is now possible to define the following transition probability:

$$\tilde{T}_K^{(l)}(x'_i, x) = \frac{\rho_K\,\left(x_{i+2^{l-1}}, x'_i; 2^{l-1}\,\Delta\,t\right)\;\rho_K\,\left(x'_i, x_{i-2^{l-1}}; 2^{l-1}\,\Delta\,t\right)}{\rho_K\,\left(x_{i+2^{l-1}}, x_{i-2^{l-1}}; 2^l\,\Delta\,t\right)}\;,$$

which is still correctly normalized (i.e. $\int\,\mathrm{d}\,x'_i\,T_K^{(l)}(x'_i, x) = 1$) because of the stability property of the Gaussian under convolution. By making use of $\tilde{T}_K^{(l)}(x'_i, x)$, it is possible to construct iteratively a so-called *Brownian bridge* between $x_{i+2^{l-1}}$ and $x_{i-2^{l-1}}$. These two extremes are kept fixed and the middle point is sampled according to $\tilde{T}_K^{(l)}(x', x)$. After that one proceeds iteratively finding the middle point between $x_{i+2^{l-1}}$ and x_i, and between x_i and $x_{i-2^{l-1}}$, employing $\tilde{T}_K^{(l-1)}(x'_{i+2^{l-2}}, x)$ and $\tilde{T}_K^{(l-1)}(x'_{i-2^{l-2}}, x)$, etc., until all the slices between are $x_{i+2^{l-1}}$ and $x_{i-2^{l-1}}$ are fixed. It is convenient to denote by $T_K^{(l,l-1)}(x'_i, x) = \tilde{T}_K^{(l-1)}(x'_{i+2^{l-2}}, x)\;\tilde{T}_K^{(l-1)}(x'_{i-2^{l-2}}, x)$, with an obvious generalization to further bisection steps.

As simple example, we work out completely the case $l = 2$:

$$\begin{aligned} T_K^{(2,2)}(x_i', x) &= \frac{\rho_K\left(x_{i+2}, x_i'; 2\,\Delta t\right)\,\rho_K\left(x_i', x_{i-2}; 2\,\Delta t\right)}{\rho_K\left(x_{i+2}, x_{i-2}; 4\,\Delta t\right)} \\ T_K^{(2,1)}(x_i', x) &= \frac{\rho_K\left(x_{i+2}, x_{i+1}'; \Delta t\right)\,\rho_K\left(x_{i+1}', x_i'; \Delta t\right)}{\rho_K\left(x_{i+2}, x_i'; 2\,\Delta t\right)} \\ &\quad \frac{\rho_K\left(x_i', x_{i-1}'; \Delta t\right)\,\rho_K\left(x_{i-1}', x_{i-2}; \Delta t\right)}{\rho_K\left(x_i', x_{i-2}; 2\,\Delta t\right)}\,, \end{aligned}$$

which exhausts all possible bisections. The product of these two terms reads:

$$\begin{aligned} T_K^{(2)}(x', x) &= T_K^{(2,2)}(x_i', x)\, T_K^{(2,1)}(x_i', x) \\ &= \frac{\rho_K\left(x_{i+2}, x_{i+1}'; \Delta t\right)\,\rho_K\left(x_{i+1}', x_i'; \Delta t\right)\,\rho_K\left(x_i', x_{i-1}'; \Delta t\right)\,\rho_K\left(x_{i-1}', x_{i-2}; \Delta t\right)}{\rho_K\left(x_{i+2}, x_{i-2}; 4\,\Delta t\right)}\,. \end{aligned}$$

The acceptance of the whole move is then given by the following equation

$$A_K(x', x) = \min\left(1,\, \frac{T_K^{(2)}(x, x_i')\, P_K(x')}{T_K^{(2)}(x_i', x)\, P_K(x)}\right) = 1\,,$$

i.e., we are still sampling exactly $P_K(x)$. Even for $T_K^{(l)}(x_i', x)$ a closed analytical expression exists:

$$\tilde{T}_K^{(l)}(x_i', x) = \left(2^l\,\pi\,\Delta t\,\Gamma\right)^{-\frac{1}{2}}\, e^{-\frac{\left(x_i' - \frac{x_{i+2^l-1} + x_{i-2^l-1}}{2}\right)^2}{2^l\,\Delta t\,\Gamma}},$$

which is a Gaussian of mean $\frac{x_{i+2^l-1} + x_{i-2^l-1}}{2}$ and variance $2^{l-1}\Delta t\,\Gamma$. The generalization of this procedure to larger values of l gives the so-called *Lévy construction* [59], which provides a fast way of simulating a free system.

C.6 The Bisection algorithm

In this section we shall describe an algorithm which is able to efficiently sample the whole $P_{eq}(x)$, by taking advantage of the Lévy construction. A new notations first. We shall employ the following slide sets:

$$\begin{aligned} I_{(l,l+1)} &= \{x_{i+2^{l-1}}, x_{i-2^{l-1}}\}, \\ I_{(l,l)} &= \{x_{i+2^{l-1}}, x_i, x_{i-2^{l-1}}\}, \\ I_{(l,l-1)} &= \{x_{i+2^{l-1}}, x_{i+2^{l-2}}, x_i, x_{i-2^{l-2}}, x_{i-2^{l-1}}\}\,, \end{aligned}$$

and so long so forth, adding middle points of the previous set at every step. We define a *partial trace* operator $\mathrm{Tr}_{(l,m)}[\cdots]$ as:

$$\mathrm{Tr}_{(l,m)}\left[\cdots\right] = \int \prod_{i \notin I_{(l,m)}} \mathrm{d}\,x_i\,[\cdots]\,.$$

With this notations, we can rewrite the transition operator of the Lévy construction step as:

$$T_K^{(l,m)}(x_i', x) = P_K^{(l,m)}(x') = \frac{\mathrm{Tr}_{(l,m)}\left[P_K(x')\right]}{\mathrm{Tr}_{(l,m+1)}\left[P_K(x')\right]} \ . \tag{C.18}$$

The probability $P_K^{(l,m)}(x)$ is well normalized and it represents a *renormalized probability* that takes care only of a subset of the whole slices involved in the construction.

Instead of accepting a move at the end of the construction of the Brownian bridge, one can accept it at every step according to the following equation:

$$A_K^{(l,m)}(x', x) = \min\left(1, \frac{T_K^{(l,m)}(x, x_i')\, P_K^{(l,m)}(x')}{T_K^{(l,m)}(x_i', x)\, P_K^{(l,m)}(x)}\right) = 1 \ .$$

The detailed balance holds at every step, as it follows from:

$$A_K^{(l,m)}(x', x)\, T_K^{(l,m)}(x_i', x)\, P_K^{(l,m)}(x) = A_K^{(l,m)}(x, x')\, T_K^{(l,m)}(x, x_i')\, P_K^{(l,m)}(x') \ ,$$

and so does for the product of all the terms involved:

$$\prod_{m=1}^{l} A_K^{(l,m)}(x', x)\, T_K^{(l,m)}(x_i', x)\, P_K^{(l,m)}(x) = \prod_{m=1}^{l} A_K^{(l,m)}(x, x')\, T_K^{(l,m)}(x, x_i')\, P_K^{(l,m)}(x') \ .$$

Rearranging this equation one can reconstruct the kinetic part of the original distribution:

$$P_K^{(l)}(x') = \prod_{m=1}^{l} P_K^{(l,m)}(x') \ ,$$

the whole transition operator – which represents the probability of generating the Brownian bridge in a single step –

$$T_K^{(l)}(x_i', x) = \prod_{m=1}^{l} T_K^{(l,m)}(x_i', x) \ ,$$

and the whole acceptance – which is the probability of accepting the whole Brownian bridge –

$$A_K^{(l)}(x', x) = \prod_{m=1}^{l} A_K^{(l,m)}(x', x) = 1 \ .$$

This notation is cumbersome for a free system, but its generalization to the interacting case leads to a simpler formulation of the so-called *bisection algorithm* [59]. The idea is simple: One propose an exact kinetic move, constructing a Brownian bridge according to the Lévy construction, and then accepts or rejects every step according to the following probability:

$$A^{(l,m)}(x', x) = \min\left(1, \frac{T_K^{(l,m)}(x, x_i')\, P_{eq}^{(l,m)}(x')}{T_K^{(l,m)}(x_i', x)\, P_{eq}^{(l,m)}(x)}\right) \ ,$$

which will be in general different from 1.

The definition of $P_{eq}^{(l,m)}(x)$ is the analog to Eq. C.18, but it is possible to employ the Trotter approximation Eq. C.5 and state that:

$$P_{eq}^{(l,m)}(x) = \begin{cases} P_K^{(l,m)}(x) \prod_{i \in I_{(l,m)}} e^{-2^{m-1}\Delta t V(x_i)} + O\left((2^{m-1}\Delta t)^3\right) & \text{PA} \\ P_K^{(l,m)}(x) \prod_{i \in I_{(l,m)}} e^{-2^{m-1}\Delta t V_{eff}(x_i)} + O\left((2^{m-1}\Delta t)^5\right) & \text{TIA} \end{cases}$$

$P_{eq}^{(l,m)}(x)$ is not normalized, but it is not a serious drawback of the approximation, since the Metropolis algorithm works well also with non-normalized distributions.

Finally we observe that the acceptance depends only on the reduced probability $\tilde{P}_{eq}^{(l,m)}(x) = \frac{P_{eq}^{(l,m)}(x)}{P_K^{(l,m)}(x)}$ which involves only the potential energy:

$$A^{(l,m)}(x', x) = \min\left(1, \frac{\tilde{P}_{eq}^{(l,m)}(x')}{\tilde{P}_{eq}^{(l,m)}(x)}\right),$$

since the kinetic part of $P_{eq}(x)$ was exactly sampled (see Eq. C.18).

The bisection scheme has the great advantage of being modular: a trial move that involves many Trotter's slices can be refused during an early bisection step, saving a lot of computation time. Unfortunately, some problems will arise from the inherent discrete nature of the algorithm. Indeed, increasing the number of bisections steps by one leads to a doubling of the slices to be moved. As we shall see in the next subsection, in order to maximize the whole acceptance of the bisection algorithm one can only tune this single discrete parameter (the bisection step), instead of a continuous one as in a classical MC (see Sec. 3.5).

C.7 Global move

The bisection scheme neglects a very natural kind of move, which is a coherent translation of all the slices (center of mass motion). If the global displacement y is distributed according to a Gaussian of zero mean and variance σ_{GM}^2, then the acceptance is given by the following equation:

$$A_{GM}(x', x) = \begin{cases} \min\left(1, \frac{\prod_{i=0}^{P-1} e^{-\Delta t V(x_i+y)}}{\prod_{i=0}^{P-1} e^{-\Delta t V(x_i)}}\right) & \text{PA} \\ \min\left(1, \frac{\prod_{i=0}^{P-1} e^{-\Delta t V_{eff}(x_i+y)}}{\prod_{i=0}^{P-1} e^{-\Delta t V_{eff}(x_i)}}\right) & \text{TIA} \end{cases}.$$

This move is very similar to a classical one, and one expects that the optimal variance σ_{GM}^2 should be correlated which the temperature $1/\beta$, as explained in Sec. 3.5.

Global moves are really effective in the last part of the annealing, when the quantum move provided by the bisection algorithm are rarely accepted. In the following sections we shall assume that a single global move is performed after every single bisection move.

C.8 Analysis of the convergence

In this section we shall present some result about the convergence of the internal energy U with respect to the number of Trotter slices employed by a PIMC algorithm with bisection moves. The data refer

to the V_{asym} double-well potential case only, because it presents a true Landau-Zener transition (see Sec. 2.2.1).

Quantum fluctuations increase the kinetic and the internal energy, but in order to have quantum effects completely at play, a suitable number of the Trotter slices must be employed. In Fig. C.2 we report the convergence to the asymptotic value of the internal energy for a $l = 1$ bisection algorithm with primitive (PA) and Takahashi-Imada (TIA) approximation. The temperature has been set to the relatively small value of $T = 0.03$ and we fixed $\Gamma = 0.5$ (i.e. an equilibrium simulation with a small mass term). As we expected from the theory, the Takahashi-Imada approximation guarantees a

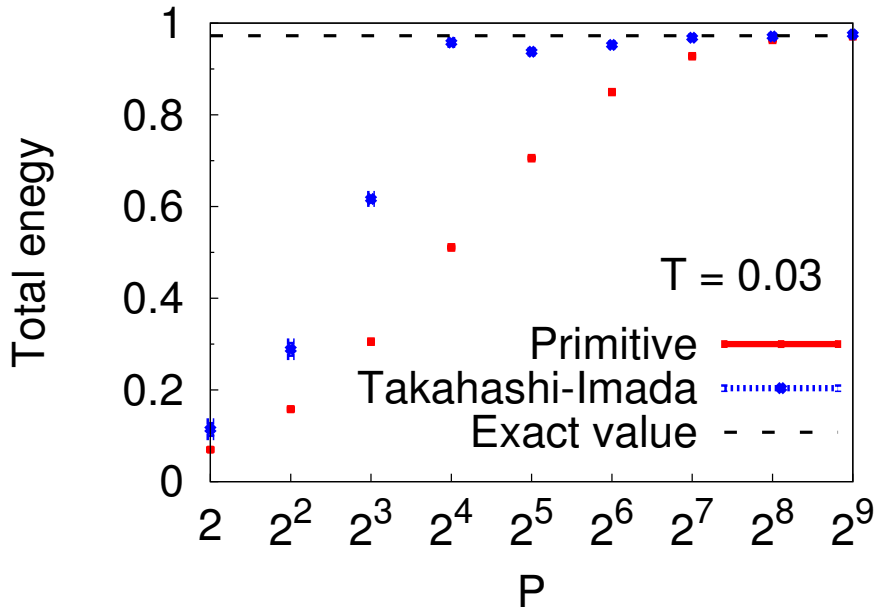

Figure C.2: Convergence plot: both curves are obtain by a level 1 bisection plus global move PIMC simulation of a particle of unitary mass ($\Gamma = 0.5$) in V_{asym} potential. A set of data refers to the primitive approximation of the action, the other to the fourth-order approximation. The exact value is obtained by means of exact diagonalization (see Chap. 2).

faster convergence (see Sec C.3). We make use of the virial estimator of the internal energy and we obtained the averages over 10^4 decorrelated configurations. [2] The larger statistical error of the first few points (small P) in the case of the fourth-order approximation is due to a smaller acceptance of the method. This is a minor drawback, and it depends on the different ratio between the potential and the kinetic part only at low P (see Eq. C.9), and on the fact that we employed exact kinetic sampling. Nevertheless, the statistical error in the convergence region agrees for both PA and TIA case.

[2] In practice, we made use of an equilibration time of 10^4 MC steps, the rest of the simulation lasted 10^8 MC steps, and we took a configuration every 10^4 MC steps. The blocking technique [52]

P	Acc	$\langle \Delta\, x^2 \rangle$
2	0.22%	$0.7912\, 10^{-3}$
4	1.5%	$0.6236\, 10^{-2}$
8	8.3%	$0.1302\, 10^{-1}$
16	38%	$0.2072\, 10^{-1}$
32	68%	$0.1425\, 10^{-1}$
64	85%	$0.5683\, 10^{-2}$
128	94%	$0.1726\, 10^{-2}$
256	97%	$0.4736\, 10^{-3}$
512	99%	$0.1275\, 10^{-3}$

Table C.1: Acceptance table of the bisection $l = 1$ algorithm with Takahashi-Imada approximation: In the second column we reported the average acceptance of a whole bisection move and in the third one we collected the average r.m.s. displacement of the single slice.

Other useful diagnostic quantities are the average acceptance of a whole bisection move and the average r.m.s. displacement $\langle \Delta\, x^2 \rangle$ of a single slice (see Sec. 3.5). In Table C.1) we report these data for the bisection $l = 1$ algorithm with Takahashi-Imada approximation. As P increases, the acceptance approaches unity, but the average r.m.s. displacement has a maximum around $P = 16$. This value is not very far from the convergence region – as shown in Fig. C.2 – and the corresponding acceptance lies within the range empirically recognized as "optimal" (around 20%) [59].

On the other hand, the same analysis applied to the primitive approximation (see Table C.2) shows that a maximum of the average r.m.s. displacement occurs when $P = 8$, which corresponds to a very poor convergence. This means that the single-step bisection algorithm is not accurate enough and a larger Brownian bridge is advisable. Increasing the number of bisection steps – the total number of Trotter's slices P being fixed – usually leads to a larger average r.m.s displacement, and also to a smaller acceptance rate. As we anticipated in the previous section, the number of bisection steps cannot be fine tuned and, as a consequence, big jumps in both average acceptance and average r.m.s. displacement are usually found.

C.9 PIMC at Landau-Zener transition

As we explained in Chap. 2 the particular form of the potential V_{asym} has been chosen because it provides a clear Landau-Zener (avoided) crossing between the ground and the first exited state. It is very instructive to study the system just at this point, which occurs for $\Gamma_c = 0.038$, corresponding to

was applied to this series in order to show that such configurations were really decorrelated.

P	Acc	$\langle \Delta x^2 \rangle$
2	3.4%	$0.1247\,10^{-1}$
4	9.9%	$0.3172\,10^{-1}$
8	24%	$0.4464\,10^{-1}$
16	47%	$0.3305\,10^{-1}$
32	70%	$0.1640\,10^{-1}$
64	86%	$0.5642\,10^{-2}$
128	94%	$0.1767\,10^{-2}$
256	97%	$0.4727\,10^{-3}$
512	99%	$0.1266\,10^{-3}$

Table C.2: Acceptance table of the bisection $l = 1$ algorithm with primitive approximation: In the second column we reported the average acceptance of a whole bisection move and in the third column we collected the average r.m.s. displacement of the single slice.

an (Hamiltonian) energy gap $\Delta\,\epsilon = 0.608\,10^{-2}$. For $T = 0.03$, this gap is the smaller energy scale in the problem and therefore thermal fluctuations could cause themselves a transition from the ground to the first exited state. Nevertheless, since the gap between the first and the second exited states at $\Gamma_c = 0.038$ is $\Delta\,\epsilon_2 = 0.498$, the system is still an effective two-level system. One can guess that at the transition point the correlation time of the algorithm is huge, because of the smallness of the gap. [3]

We begin by examining the convergence plot of a $l = 1$ and $l = 5$ bisection algorithm at $T = 0.03$. We made use of the virial estimator for the internal energy and we obtained the observables values averaging over 10^3 decorrelated configurations. [4] Fig. C.3 shows that, despite the fact that the convergence seems to be reached when $P \geq 40$ in both cases, the $l = 1$ algorithm displays a less regular behavior. We introduce now a new test, particularly useful in such a correlated system. In Fig. C.4 we report the centroid coordinate of the configurations we employed in order to find the averages. The first panel of the figure clearly shows a lack of ergodicity of the $l = 1$ bisection algorithm. The centroid is prone to be stack in one of the two minima (i.e. $x_- \simeq -0.75$ and $x_+ \simeq +1.25$ see Sec. 2.2), and, indeed, it presents many "holes" in the centroids data. On the other hand the $l = 5$ algorithm display a "denser" distribution, a fingerprint of a truly ergodic behavior.

[3] This correlation time has nothing to do with characteristic time $\tau_{LZ} \propto \Delta\,\epsilon^{-2}$ of an Landau-Zener transition. It means only that two configurations obtained by means of an *equilibrium* simulation at $\Gamma = \Gamma_c$ should be decorrelated only if they were separated by an amount of MC steps greater than this correlation time.

[4] In particular we performed 10^9 MC steps employing an equilibration time of 10^6 MC steps (i.e. we took a configuration every 10^6 MC steps). These times are considerably longer than for the case $\Gamma = 0.5$ we studied above.

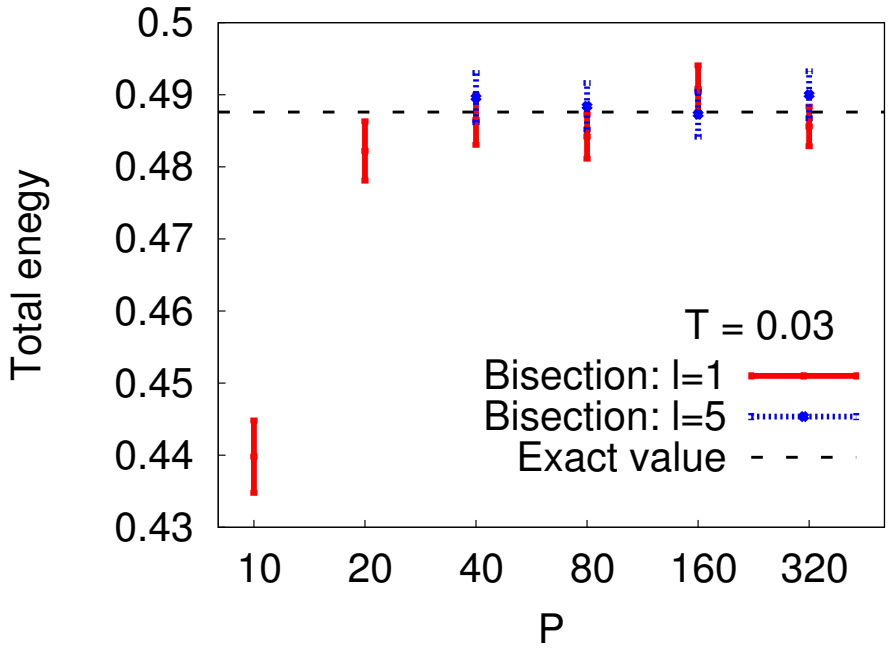

Figure C.3: Convergence plot of the $l = 1$ and $l = 5$ bisection algorithm at the Landau-Zener transition point: V_{asym} potential case. Data referring to the $l = 5$ case are computed only for $P \geq 2^5 = 32$ due to the constraint in the path construction. The exact vale is obtained by means of exact diagonalization (see Sec. 2.2).

Summarizing, it is better, in terms of computational time, to avoid exceedingly large correlation times, by increasing the number of bisection steps, rather than having to increase the number of iterations with a small bisection level.

C.10 The role of the initial conditions

We would like to spend here some words about the initial condition we implemented in the actual annealing simulations. As we explained in Sec. 3.5, every simulation starts with an equilibration part, consisting of a few (with respect to the whole simulations length) steps performed at fixed annealing parameter Γ. We found that $\Gamma = 0.5$ is a reasonable value, as one can see in Fig. C.5, and also because of the choice in Sec. 2.2. It remains to choose a suitable number of Trotter's slices and of bisection steps.

In Fig. C.5 we plot the profile of the diagonal part of the density matrix at $T = 0.03$ and $\Gamma = 0.5$. For every coordinate x this observable is equal to the average $\langle \chi_{[x-\mathrm{d}x, x+\mathrm{d}x]} \rangle$, where $\mathrm{d}x$ depends on the

discretization grid employed. [5] Its exact value was obtained by diagonalization of the corresponding Hamiltonian, taking the usual thermal average over the low lying eigenvectors. In order to maintain each simulation around the optimal sampling regime, we always made use of a number bisection steps which maximizes the average r.m.s. displacement. In particular we employed $l = 2$ for $N_T = 20$, $l = 3$ for $N_T = 40$, $l = 4$ for $N_T = 80$, and $l = 5$ for $N_T = 160$. The exponential scaling is evident. Indeed, for every practice purpose, the ratio between the moved slices, $2^l - 1$, and P must remain constant in order to have optimal sampling conditions.

We want to stress that a sub-optimal choice of P leads to wrong initial condition. The exact solution at $T = 0.03$ and $\Gamma = 0.5$ is a unimodal distribution, meaning that the barrier among the well does not affect the system. However, for $P < 160$, the diagonal part of the density matrix presents two peaks, more or less defined. One can also claim that, for sub-optimal number of Trotter's slice, the system is less "quantum", due to small tunneling rate.

Unfortunately a simulation with $P = 160$ requires a lot of computational time, even according to a $l = 5$ bisection scheme. However, we stress that, for small values of Γ (i.e. for large "mass"), this initial large amount of Trotter's slices is likely to be inessential. Indeed in such parameter region the system is "less quantum" due to its larger mass. In order to check this hypothesis, we performed a series of Quantum Annealings, starting with different number of Trotter slices P. In Fig. C.6 we plot the residual potential energy ϵ_{res} versus the annealing time τ for all the cases considered. Despite the fact that, initially, the systematic error due to a sub-optimal Trotter's slices number is considerable, it becomes less and less relevant for slower annealings (i.e. larger τ).

These numerical experiments suggest that the actual asymptotic annealing behavior can be effectively studied even by means of a simulation with $P = 20$, saving a lot of computational time.

C.11 The Instanton move

Even though the PIMC algorithm presents many drawbacks (finite temperature T, sampling problems for the action, difficulties with the Trotter break-up), we can still manage in order to find a smarter proposal move. In particular we try to add a very specific global – but quantum – displacement of the whole PIMC configuration, taking advantage from the so-called *instanton theory*.

This theory applies to a kind of potential more symmetric than the ones that we considered so far (see Sec. 2.2). It reads:

$$V(x) = V_0 \, (x^2 - a^2)^4. \tag{C.19}$$

The classical equation of motion in such a potential is given by:

$$\frac{1}{2\,\Gamma}\,\ddot{x} = -\,\frac{\partial V}{\partial\, x}\ ,$$

but it is better to study its analytic continuation (i.e. setting $t = -i\,\tau$):

$$\frac{1}{2\,\Gamma}\,\frac{\partial^2 x}{\partial\,\tau^2} = \frac{\partial V}{\partial\, x}\ . \tag{C.20}$$

[5] In this case the mesh is $\mathrm{d}x = 0.05$. The error is obtained over a set of 10^3 decorrelated configurations.

The trivial solutions of the last equation are $x(\tau) = \pm a$. Moreover, other trajectories in the imaginary time – which are also solutions of Eq. C.20 – can be written as:

$$x_{cl}(\tau) = \pm a \, \tanh\left(\frac{\omega\,(\tau - \tau_0)}{2}\right) \quad \text{where} \quad \omega^2 = \frac{8\,\Gamma\,V_0}{a^2} \, .$$

They are the so-called the *instantonic* solutions. [6]

These paths connect the two minima, starting from $\mp a$ at $\tau = -\infty$ and reaching $\pm a$ at $\tau = +\infty$. It is also possible to estimate the (imaginary) time which a particle takes tunneling through the barrier. It reads:

$$\tau_{tun} \simeq \frac{1}{\omega} = \sqrt{\frac{a^2}{16\,\Gamma\,V_0}} \, .$$

We stress that – since τ_{tun} depends on Γ – this time will be very large in the final part of an annealing simulation.

The (anti)instanton is not by itself a good PIMC configuration, because it is not a closed path in the configuration space. In order to obtain a suitable path, one has to consider a linear superposition of an instanton and an antiinstanton. This configuration reads:

$$x_i^{(ins)} = \pm \left[a \, \tanh\left(\frac{\tilde{\omega}\,(i - i_1)}{2}\right) - a \, \tanh\left(\frac{\tilde{\omega}\,(i_2 - i)}{2}\right)\right] \, , \tag{C.21}$$

where the sign and the constants i_1 and i_2 are free parameters, while $\tilde{\omega} = \omega\,\Delta\,t$. Strictly speaking, Eq. C.21 does not represent an actual periodic solution, because in general $x_{P+i}^{(ins)} \neq x_i^{(ins)}$. However, till $\frac{1}{\tilde{\omega}} < \frac{P}{2}$, the error is negligible. Anyway, for small Γ the instanton-like displacement will not available, since the above condition can not be longer satisfied.

In a practical implementation of the PIMC, we propose an instanton move, Eq. C.21, after any bisection (and global) move. Our algorithm will randomly choose the sign and the two parameters i_1 and i_2 of the Eq. C.21, while the coefficient a is set to 1, since $2a \simeq a_+ - a_- = 2$ in the case our double-well potentials (see Sec. 2.2). After that $x_i^{(ins)}$ has been generated, a proposal configuration is obtained by adding it to the older configuration: $x_i = x_i^{(old)} + x_i^{(ins)}$, and it is accepted or rejected according to the usual Metropolis algorithm. In this case acceptance operator simply reads:

$$A_{ins}(x', x) = \min\left(1, \frac{P_{eq}(x')}{P_{eq}(x)}\right) \, ,$$

where $x = x^{(old)}$ and $x' = x^{(old)} + x^{(ins)}$.

C.12 The Lorentzian Move

In Sec. 4.2 we present some simulations performed according to the Hamiltonian:

$$H = \Gamma(t)\,p^2 + V(x) \, ,$$

[6] By convention we shall call *instanton* the solution beginning with a + sign and *antiistanton* the other one.

where $p = -i\,\frac{\partial}{\partial x}$ is the momentum operator. This is a natural choice, but it is not the only one.

Given insight from the *relativistic dynamics*, one can set a new Hamiltonian as:

$$H = \Gamma(t)\,|p| + V(x)\ .$$

The new kinetic operator $\Gamma(t)\,|p|$ is singular in real space representation, but the correspondent kinetic density operator can be written in a simple closed form:

$$\rho_K\,(x', x; \Delta\,t) = \frac{1}{\pi}\,\frac{\Gamma\,\Delta\,t}{\Gamma^2\,\Delta\,t^2 + (x'-x)^2} = \frac{1}{\pi\,\Gamma\,\Delta\,t}\,e^{-\ln\left(1+\left(\frac{x'-x}{\Gamma\,\Delta\,t}\right)^2\right)}\ .$$

This is a *Lorentzian (or Cauchy) distribution*, while the kinetic density operator, Eq. C.16, considered so far was a Gaussian one.

Since in the Sec. 3.5 we realized that a proposal move distribution with long tails can improve the Classical Annealing performance, we want now to test this hypothesis also for the PIMC-QA dynamics. It is possible to obtain a generalization of the Lévy construction for this kind of Lorentzian move, observing that both Gaussian and Lorentzian distribution are stable under convolution (a property shared by all the so-called *Lévy distribution* [92, 93]).

First of all, we have to find a closed expression of $T_K^{(1)}(x_i', x)$, as defined in Eq. C.17. In the Lorentzian case it reads:

$$T_K^{(1)}(x_i', x) = \frac{\Gamma\,\Delta\,t}{2\,\pi}\,\frac{\left(\frac{x_{i+1}-x_{i-1}}{\Gamma\,\Delta\,t}\right)^2 + 4}{\left[\left(\frac{x_{i+1}-x_i'}{\Gamma\,\Delta\,t}\right)^2 + 1\right]\,\left[\left(\frac{x_i'-x_{i-1}}{\Gamma\,\Delta\,t}\right)^2 + 1\right]}\ .$$

Unfortunately, this is no longer a Lorentzian distribution, and therefore there is not an obvious sampling way (that, for instance, makes use of the usual inversion technique; see [94][chap. 33]).

Carry the algebra on, after a smart change of variable (and taking care of the right Jacobian of such a transformation), we obtain a simpler form of the transition probability. It is given by:

$$T_K'^{(1)}(y, a) = \frac{2}{\pi}\,\frac{a^2+1}{\left[(y+a)^2+1\right]\,\left[(y-a)^2+1\right]}\ ,$$

where

$$\begin{aligned} a &= \frac{x_{i+1} - x_{i-1}}{2\,\Gamma\,\Delta\,t} \\ c &= \frac{x_{i+1} + x_{i-1}}{2\,\Gamma\,\Delta\,t} \\ y &= \frac{x_i'}{\Gamma\,\Delta\,t} - c\ . \end{aligned}$$

Finally (after some more tedious algebra), we find that:

$$T_K'^{(1)}(y, a) = \underbrace{\frac{1}{\pi}\,\frac{y^2+a^2+1}{\left[(y+a)^2+1\right]\,\left[(y-a)^2+1\right]}}_{W_1(y,a)} - \underbrace{\frac{1}{\pi}\,\frac{y^2-a^2+1}{\left[(y+a)^2+1\right]\,\left[(y-a)^2+1\right]}}_{W_2(y,a)}\ ,$$

and also that $T_k^{\prime(1)}(y,a) < 2\,W_1(y,a)$.

An efficient strategy to sample $T_k^{\prime(1)}(y,a)$ is to sample before $W_1(y,a)$ and then to make use of the usual rejection technique (see [94][chap. 33]) in order to achieve the original distribution. The primitive of $W_1(y,a)$ can be computed and it reads:

$$M_1(u,a) = \int_{-\infty}^{u} \mathrm{d}\,y\, W_1(y,a) = \frac{1}{2\,\pi}\,\tan^{-1}\left(\frac{2\,u}{1-u^2+a^2}\right) + \frac{1}{2}\;.$$

Inverting $M_1(u,a)$, and making use of an equidistributed random number $u \in (0,1]$, we find that:

$$y = \begin{cases} -d - \sqrt{d^2+(a^2+1)} & u \in (0,\frac{1}{2}] \\ -d + \sqrt{d^2+(a^2+1)} & u \in (\frac{1}{2},1] \end{cases} \qquad \text{where} \quad d = \cot\left[2\,\pi\,\left(u-\frac{1}{2}\right)\right]\;.$$

Provided $u' \in (0,1]$ – another equidistributed random number – we shall accept the y obtained above according to the condition:

$$2\,W_1(y,a)\,u' \leq W_1(y,a) - W_2(y,a)\;.$$

It results from standard considerations (see [94][chap. 33]) that the average acceptance of this method will be around 50%. It means that in order to generate N random numbers distributed according to $T_K^{(1)}(x_i',x)$, one has to try on average $2\,N$ times. Finally the original middle point value x_i' appearing in $T_K^{(1)}(x_i',x)$ is obtained inverting the early change of variables:

$$x_i' = \Gamma\,\Delta\,t\;\;(y+c)\;\;.$$

This sampling method is a bit cumbersome, but it can be easily implemented as a computer algorithm. Its generalization to the more general case, namely $T_K^{(l)}(x_i',x)$, can be easily obtained following the recipes of App. C.5.

Since we proved that it is possible to implement a bisection algorithm for a relativistic Hamiltonian, one can also proceed as in Sec. C.6 and implement a real Lorentzian-PIMC. Moreover, since $[V,[|p|,V]] = 0$, for such a Lorentzian-PIMC the simple primitive approximation is already exact till the fourth-order! We finally recall that the virial estimators for the kinetic an potential energy simply read:

$$\begin{aligned} \tilde{K}_{PA} &= \frac{1}{\beta} + \frac{1}{P}\sum_{i=0}^{P-1}(x_i-\bar{x})\,\frac{\partial\,V(x_i)}{\partial\,x_i} \\ \tilde{V}_{PA} &= \frac{1}{P}\sum_{i=1}^{P} V(x_i)\;, \end{aligned}$$

where $\bar{x}$ is the centroid coordinates (see Sec. C.3).

A further generalization of the Lorentzian-PIMC, suitable for higher dimensional configuration space, is discussed in Sec. 4.5.

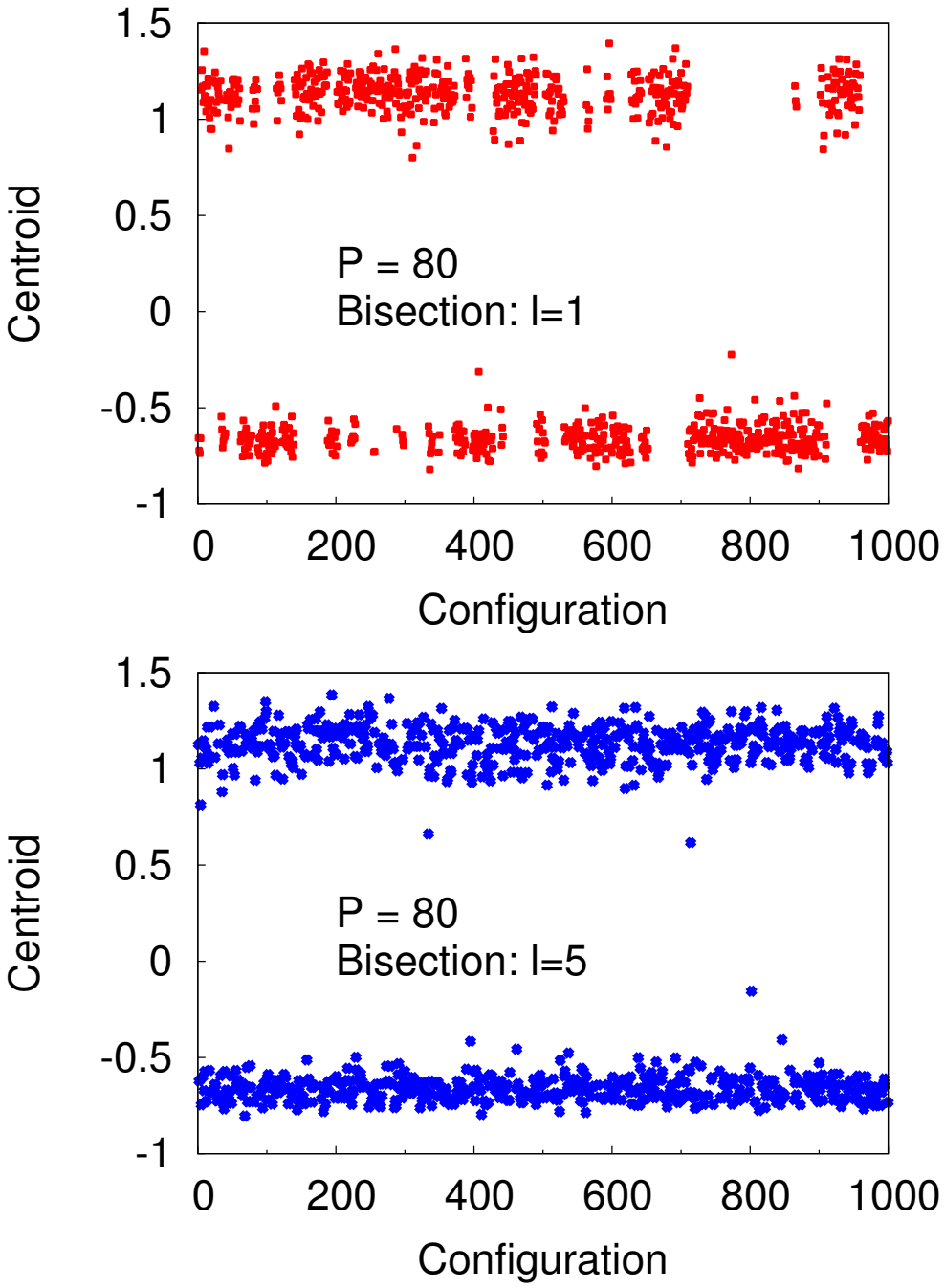

Figure C.4: Centroid coordinate of the $l = 1$ (first panel) and $l = 5$ (section panel) bisection algorithm at $\Gamma = \Gamma_c$ and for V_{asym} potential. These data are obtained using $P = 80$ Trotter's slices. We only reported the centroid coordinate of the configurations involved in the averages (see the text).

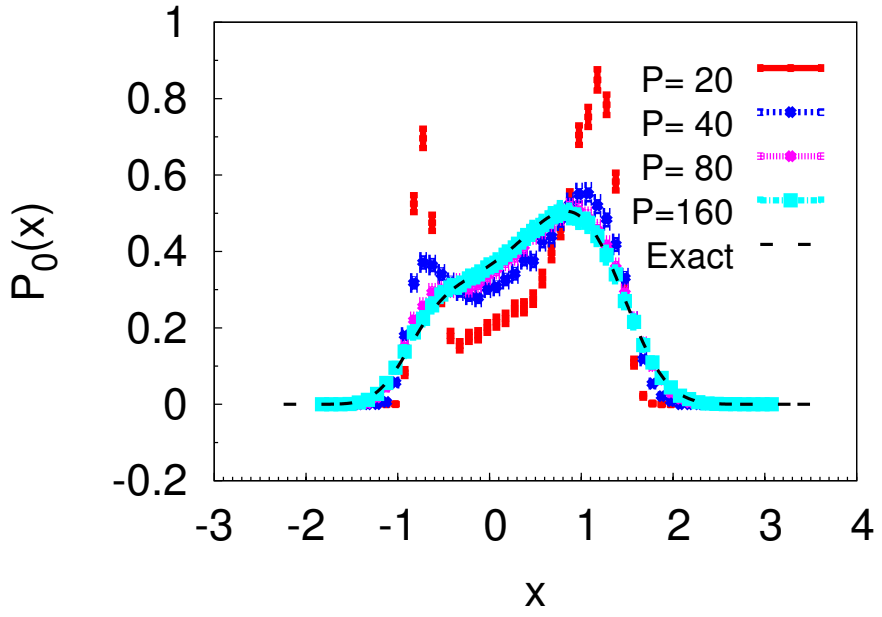

Figure C.5: Convergence of the ground state for the initial condition $\Gamma = 0.5$. These data refers to the case of V_{asym} potential.

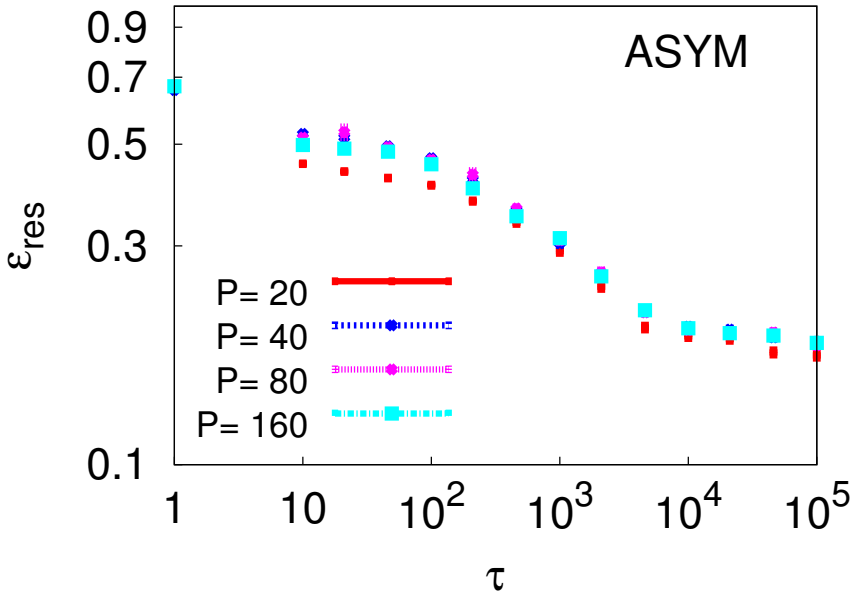

Figure C.6: Quantum annealing in V_{asym} potential: Different lines correspond to different number of Trotter slices. The correspondent number of bisection steps is discussed in the text.

Appendix D

Appendix of Chapter 5

D.1 More about GFMC

The basic idea of the Green's function Monte Carlo (GFMC), Diffusion Monte Carlo (DMC) and other similar *projection techniques* is to obtain the exact ground state $\psi_0(x)$ of an Hamiltonian operator H from a simple trial function $\psi_T(x)$, by iterating the application of a (simple) function of the Hamiltonian $G(H)$[95]:

$$\lim_{n\to\infty} G^n(H)\,\psi_T = \psi_0\ . \tag{D.1}$$

It's customary to use only the following cases:

$$G(H) = \begin{cases} e^{-\Delta t\,(H-\epsilon_T)} & \text{DMC} \\ \frac{1}{1+\Delta t\,(H-\epsilon_T)} & \text{GFMC} \end{cases}\ ,$$

where ϵ_T is the so-called *trial energy* – an estimate of the true ground state energy – and Δt is a constant governing the projection rate (e.g. $\Delta t = 0$ means no projection at all). We recall that we set $\hbar = 1$ throughout all this appendix.

To prove the converge of the algorithms, Eq D.1, we shall make use of the eigenvectors $\{\psi_j(x)\}$ and the eigenvalues $\{\epsilon_j\}$ of the Hamiltonian. In such a basis the trial function reads: $\psi_T(x) = \sum_{j\geq 0} a_j\,\psi_j(x)$ (the coefficients a_j account for the initial condition). Therefore the projection process is given by:

$$\lim_{n\to\infty} G^n(H)\,\psi_T = \lim_{n\to\infty} \begin{cases} \sum_{j\geq 0} e^{-n\,\Delta t\,(\epsilon_j-\epsilon_T)}\,a_j\,\psi_j(x) & \text{DMC} \\ \sum_{j\geq 0} \left(\frac{1}{1+\Delta t\,(\epsilon_j-\epsilon_T)}\right)^n a_j\,\psi_j(x) & \text{GFMC} \end{cases}\ .$$

As a consequence – whenever there is a finite gap (i.e. $\epsilon_1 - \epsilon_0 > 0$), provided an appropriate choice of Δt and ϵ_T, and also that ψ_0 and ψ_t are non-orthogonal (i.e. $a_0 \neq 0$) – the algorithms will converge to the actual ground state, apart for an immaterial multiplicative constant.

In the case of lattice system, it is more convenient to deal with a short-time approximation of the aforementioned cases, which is given by:

$$G(H)(x', x) = 1 - \Delta t\,(H(x', x) - \epsilon_T)\ ,$$

where $H(x',x)$ stands for $\langle x'|H|x\rangle$, the matrix elements between the two configurations x', x. We emphasize that the matrix representation of $G(H)$ is particularly easy to find in the case of the Random Ising Hamiltonian defined in Eq. 5.1, since it connects only nearest-neighbors.

The trial function $\psi_T(x)$ is supposed to be a simple analytical many-body function, which is also a good approximation of the actual ground state. A huge amount of methods have been developed in order to construct such a mathematical object [70, 72, 96] – at least in the context of the electronic structure – and we have discussed a couple of especially simple choices of $\psi_T(x)$ in Chap. 5.

Whatever trial function is employed, one can form another useful operator starting from the original projector, $G(H)$:

$$\bar{G}(H)(x',x) = \psi_T(x')\, G(H)(x',x)\, \frac{1}{\psi_T(x)} \; .$$

We note that the iterated application of this function yields to $\psi_0(x)\,\psi_T(x)$ instead that on ground state $\psi_0(x)$. Nevertheless – as we shall see just below – this choice guarantees better numerical stability of the algorithm.

Unfortunately neither $\bar{G}(H)$ nor $G(H)$ are native stochastic matrices – their elements can be negative and their column can add to constants different from 1 (i.e. $\sum_{x'} G(H)(x',x) \neq 1$) – which is basis condition for the transition operator of a Markov chain (see. Sec. 3.1). Nevertheless it is still possible to manage these shortcomings and make a stochastic implementation of the projection algorithm, Eq. D.1, available. First of all, both $G(H)$ and $\bar{G}(H)$ are operators with non-negative entries whenever $\psi_T(x) > 0$, $H(x',x) \leq 0$ if x', x, and $\Delta\, t$ is small enough (we stress that – using $\psi_T(x) = 1$ – the projector $G(H)$ can be obtained from $\bar{G}(H)$ as a special case). This is far to be the general rule and in many cases there is no way to reduce $G(H)$ to a non-negative form (this is the so-called *sign problem* of the Quantum Monte Carlo [51]).

If it was been possible to find such a non-negative projector, it would be simple to derive a good transition operator starting, for instance, from $\bar{G}(H)$. One first introduce the following column normalization factor:

$$b(x) = \sum_{x'} \bar{G}(H)(x',x) \; ,$$

($\sum_{x'}$ is understood as the appropriate integral in the case of a continuous configuration space) and then it is possible to define

$$T(x',x;\Delta\, t) = \frac{\bar{G}(H)(x',x)}{b(x)} \; , \tag{D.2}$$

which turns out to be a real stochastic operator. Moreover the normalization factor, $b(x)$, can be rewritten by means of the so-called *local energy*, $\epsilon_L(x) = \sum_{x'} \frac{\psi_T(x')}{\psi_T(x)} H(x',x)$, as follow:

$$b(x) = 1 - \Delta\, t\; (\epsilon_L(x) - \epsilon_T) \; .$$

One can see from the definition that, if $\psi_T(x) = \psi_0(x)$, then the local energy will be equals to the ground state energy, $\epsilon_L(x) = \epsilon_0$. Therefore, the better is the local energy, the smaller will be the fluctuations induced by the normalization factor on the whole algorithm

Unfortunately, this Markov chain it is not yet equivalent to the projection process defined in Eq. D.1, but it is a fundamental ingredient of a true implementation. Therefore it is worth to take

just a bit more space in order to develop some fundamental concept about such a Markov chain that we shall employ after.

Carrying on the previous algebra, we found that the transition probability up to the first order in the time interval $\Delta\, t$ reads:

$$T(x', x; \Delta\, t) = \begin{cases} -\Delta\, t\, \frac{\psi_T(x')}{\psi_T(x)}\, H(x', x) + O(\Delta\, t^2) & \text{if} \quad x' \neq x \\ 1 - \Delta\, t\, (H(x, x) - \epsilon_T) + O(\Delta\, t^2) & \text{if} \quad x = x \end{cases} .$$

As a consequence, by using the above definition of the transition operator, it is also possible to define a *continuous-time* Markov process based on the projection function $G(H)$ whose transition rate from the configuration x to configuration x' is given by:

$$T(x', x) = \lim_{\Delta\, t \to 0} \frac{T(x', x; \Delta\, t)}{\Delta\, t} = -\frac{\psi_T(x')}{\psi_T(x)}\, H(x', x) \quad \text{if} \quad x' \neq x\ . \tag{D.3}$$

Since the *escaping probability* from a configuration x (in an infinitesimal time interval $\Delta\, t$) is naturally defined as:

$$p_{esc}(x; \Delta\, t) = 1 - T(x, x; \Delta\, t) = \sum_{x' \neq x} T(x', x; \Delta\, t) = \Delta\, t\ (H(x, x) - \epsilon_L(x)) + O(\Delta\, t^2)\ ,$$

the corresponding *escaping rate* reads:

$$\lambda_x = \lim_{\Delta\, t \to 0} \frac{p_{esc}(x; \Delta\, t)}{\Delta\, t} = H(x, x) - \epsilon_L(x)\ .$$

We stress that such an escaping rate is non-negative because of the inequality $\epsilon_L(x) = H(x, x) + \sum_{x' \neq x} \frac{\psi_T(x')}{\psi_T(x)}\, H(x', x)$ (remember that $H(x', x) \leq 0$ if $x' \neq x$ and that $\psi_T(x) > 0$). Finally, the escaping probability from the configuration x in a *finite* time, $t = n\, \Delta\, t$, can be extrapolated as:

$$P_x(t) = \lim_{n \to \infty} (1 - \Delta\, t\, \lambda_x)^{n-1}\ \lambda_x = e^{-t\, \lambda_x}\ \lambda_x\ . \tag{D.4}$$

We briefly stop here the discussion and we summarize the way the continuous-time Markov chain (or kinetic Monte Carlo [97]) proceeds:

1. First extract an equidistributed random number $u_1 \in (0, 1]$ and let the time counter to advance by an amount equals to:
$$t_{incr} = -\frac{\ln u_1}{\lambda_x}\ . \tag{D.5}$$
This is equivalent to the Poisson process defined in the Eq. D.4 [98].

2. Given an enumeration of all the possible final configurations $\{x'_i\}_{i \in I}$ reachable from x, extract another equidistributed random number $u_2 \in (0, 1]$ in order to find the new configuration x_i sampling the conditional probability: $P(x_i|x) = \frac{T(x_i, x)}{\sum_i T(x_i, x)}$ [98]. The method is straightforward whenever only a finite number of final configurations is available $I = \{1, \ldots, N_x\}$ (e.g. in the case of a Random Ising Model). It suffices, indeed, to fulfill the condition:
$$\sum_1^{i-1} T(x'_j, x) \leq u_2 \left(\sum_j T(x'_j, x) \right) < \sum_{j=1}^{i} T(x'_j, x)\ , \tag{D.6}$$
which can be easily obtained by means of a standard computer routine [74].

We emphasize that one may save a lot of CPU time by using such a continuous-time sampling. Otherwise – by means of standard Metropolis' algorithm (see Sec. 3.2) – one could propose a random move and accept it according to the transition probability defined in Eq. D.2 every time step, Δt. Unfortunately, if the transition probability is really small, one can easily get stuck for a lot of (Monte Carlo) steps in the same configuration (i.e. for a time which is a large multiple of Δt, let us say $t = n\,\Delta t$). By means of the continuous-time algorithm, instead, every move is always accepted, and it suffice to advance the internal time by $t = n\,\Delta t$. The trade-off is that the internal time increasing t is no more a parameter (i.e. always equals to Δt), but a true a random variable.

As we said above, such a Markov chain (even in the continuous-time fashion) does not implement the right projection process we need. In particular it converges to a probability distribution $\psi_T^2(x)$, instead of the ground state $\psi_0^2(x)$. To prove this statement it suffices to show that the detailed balance condition holds:

$$T(x', x)\,\psi_T^2(x) = T(x, x')\,\psi_T^2(x') \ ,$$

and it can be done in few lines. As a consequence, the continuous-time Markov chain we define above is just a smart way to implement the so-called *Variational Monte Carlo* [51, 72]. In particular, it allows to evaluate the observable averages on the trial function, $\langle\psi_T|O|\psi_T\rangle$ and – whether ψ_T depends on one or more external parameters – it can be used in order to find their optimal optimal value (i.e. the set of parameter that provides better estimate of some observable, usually the total energy, $\langle\psi_T|H|\psi_T\rangle$). This can be achieved in several clever way [70, 96] or simply by inspection, as we did in Chap. 5. It is worth to note that one can also deal with many walkers (i.e. many independent Markov chains) at time as a simple way to increment the accumulated statistics. In the rest of the appendix we shall refer to the averages obtained employing a such continuous-time Markov chain (single- or multi-walker) as the *unweighted averages* (or variational averages). They will be indicated by the symbol $\langle\cdots\rangle_T$. For instance, the expectation value of the Hamiltonian over the trial function $\psi_T(x)$ (the so-called *variational (total) energy*) is given by the unweighted average of the local energy $\epsilon_L(x)$

$$\epsilon_T = \langle\psi_T|H|\psi_T\rangle = \int \mathrm{d}x\, \frac{\langle\psi_T|H|x\rangle}{\langle\psi_T|x\rangle}\,\psi_T^2(x) = \frac{\sum_{j=1}^{M}\langle\epsilon_L^{(j)}\rangle_T}{M} \ , \tag{D.7}$$

where M is the walker number (the so-called *population* and $j \in \{1, \dots M\}$ the single walker index.

We emphasize that the correlation time of a continuous-time Markov chain implemented by means of $G(H)$ can be obtained by means of the spectral analysis explained in Chap. 3.6 (for a single walker). In particular the symmetrization of the transition operator is the original projector (see Eq. D.9), whatever trial function is employed (it can be proved simply by means of the definition of $\bar{G}(H)$). Therefore the correlation time of the variational approach is always proportional to the inverse of the original Hamiltonian (first) gap: $t_{corr} \propto \frac{1}{\epsilon_1 - \epsilon_0}$. This is also true for a multiwalker scheme, since the walkers dynamics is completely independent.

Now we want to face the problem to find a real stochastic implementation of the projection algorithm defined in Eq. D.1. In order to obtain the projection of the trial function on the ground state one must extend the configuration space adding a new degree of freedom to every walker: the *weight* w. As a consequence, the whole walker configuration reads now: $(x^{(j)}, w^{(j)})$, where $j \in \{1, \dots M\}$ is

the walker index. For the sake of simplicity we drop the walker index till we shall really need of it (i.e. we shall consider first the single walker formalism).

We assume that transition rate from (x, w) to (x', w') is given by:

$$T(x', w', x, w; \Delta t) = T(x', x; \Delta t)\, \delta\left(\frac{w'}{b(x)} - w\right) \frac{1}{b(x)} \,. \tag{D.8}$$

As a consequence the weight updating is obtained through a *multiplicative process* (i.e. $w' = w\, b(x)$) implemented over the usual Markov chain. Suppose further that in the extended space the initial condition is given by the probability: $P_0(x, w) = \psi_T^2(x)\, \delta(w - 1)$. It immediately follows that the wavefuntion is given by:

$$\frac{1}{\psi_T(x)} \int \mathrm{d}w\, w\, P_0(x, w) = \psi_T(x).$$

More in general we want that:

$$\psi_{(n)}(x) = \frac{1}{\psi_T(x)} \int \mathrm{d}w\, w\ P_{(n)}(x, w) \,,$$

where

$$\begin{aligned} P_{(n)}(x', w') &= \int \mathrm{d}w \sum_{x'} T(x', w', x, w; \Delta t)\, P_{(n-1)}(x, w) \\ \psi_{(n)}(x') &= \sum_{x} G(H)(x', x; \Delta t) \psi_{(n-1)}(x) \,, \end{aligned}$$

(and $\psi_{(0)}(0) = \psi_T(x)$, by consistence). These equations define an isomorphism between the Markov process in the extended space and the projection of the trial function on the ground state. To prove this fact we note that:

$$\begin{aligned} \psi_{(n)}(x') &= \langle x'|G(H)|\psi_T\rangle \\ &= \frac{1}{\psi_T(x')} \int \mathrm{d}w'\, w'\, P_{(n)}(x', w') \\ &= \frac{1}{\psi_T(x')} \int \mathrm{d}w'\, w' \int \mathrm{d}w \sum_{x} T(x', w', x, w; \Delta t)\, P_{(n-1)}(x', w') \\ &= \frac{1}{\psi_T(x')} \int \mathrm{d}w'\, w' \int \mathrm{d}w \sum_{x} T(x', x; \Delta t)\, \delta\left(\frac{w'}{b(x)} - w\right) \frac{1}{b(x)} P_{(n-1)}(x', w') \\ &= \frac{1}{\psi_T(x')} \int \mathrm{d}\, w \sum_{x} b(x)\, w\, T(x', x; \Delta t)\ P_{(n-1)}(x', w') \\ &= \sum_{x} \frac{1}{\psi_T(x')} \bar{G}(H)(x', x; \Delta t) \int \mathrm{d}w\, w\,, P_{(n-1)}(x', w') \\ &= \sum_{x} G(H)(x', x; \Delta t)\, \frac{1}{\psi_T(x)} \int \mathrm{d}w\, w\,, P_{(n-1)}(x', w') \\ &= \sum_{x} G(H)(x', x; \Delta t)\, \psi_{(n-1)}(x) \end{aligned}$$

Therefore the following equation holds:

$$\langle x'|G(H)|\psi_T\rangle = \frac{1}{\psi_T(x')} \int \mathrm{d}\, w'\, w' \int \mathrm{d}\, w \sum_{x} T(x', w', x, w; \Delta t)\, \psi_T^2(x)\, \delta(w - 1)$$

$$= \frac{1}{\psi_T(x')} \sum_x b(x)\, T(x', x; \Delta t)\, \psi_T^2(x) \ ,$$

and – more in general – it follows that:

$$\langle x'|G(H)^n|\psi_T\rangle = \frac{1}{\psi_T(x')} \sum_{x_{n-1},\ldots,x_1,x_0} w(x', x_{n-1}, \ldots, x_1, x_0)\, T(x', x_{n-1}, \ldots, x_1, x_0) \ , \quad \text{(D.9)}$$

where

$$w(x', x_{n-1}, \ldots, x_1, x_0) = \prod_{i=0}^{n-1} b(x_i) \ , \quad \text{(D.10)}$$

is the weight cumulated during the stochastic process $(x', x_{n-1}, \ldots, x_1, x_0)$. The conditional probability to follow that path in the configuration space is clearly given by $T(x', x_{n-1}, \ldots, x_1, x_0) = \prod_{i=0}^{n-1} T(x_{i+1}, x_i; \Delta t)\, \psi_T^2(x_0)$, with $x' = x_n$. Finally the average of the energy can be attained by means of the following limit:

$$\begin{aligned} \frac{\langle\psi_0|H|\psi_0\rangle}{\langle\psi_0|\psi_0\rangle} &= \lim_{n\to\infty} \frac{\langle\psi_T\, G(H)^{\frac{n}{2}}|H|G(H)^{\frac{n}{2}}\,\psi_T\rangle}{\langle\psi_T\, G(H)^{\frac{n}{2}}|G(H)^{\frac{n}{2}}\,\psi_T\rangle} \\ &= \lim_{n\to\infty} \frac{\langle\psi_T|H|G(H)^n\,\psi_T\rangle}{\langle\psi_T|G(H)^n\,\psi_T\rangle} \\ &= \int \mathrm{d}x\, \epsilon_L(x)\, \Pi_{eq}(x) \ , \end{aligned}$$

where

$$\begin{aligned} \Pi_{eq}(x) &= \frac{\langle\psi_T|x\rangle\,\langle x|\psi_0\rangle}{\langle\psi_T|\psi_0\rangle} \\ &= \lim_{n\to\infty} \frac{\langle\psi_T|x\rangle\,\langle x|G(H)^n\,\psi_T\rangle}{\langle\psi_T|G(H)^n\,\psi_T\rangle} \\ &= \frac{\sum_{x_{n-1},\ldots,x_1,x_0} w(x, x_{n-1}, \ldots, x_1, x_0)\, T(x, x_{n-1}, \ldots, x_1, x_0)}{\sum_{x_n,\ldots,x_1,x_0} w(x_n, x_{n-1}, \ldots, x_1, x_0)\, T(x_n, x_{n-1}, \ldots, x_1, x_0)} \\ &= \frac{\langle\delta(x - x_n)\, w(x_n, x_{n-1}, \ldots, x_1, x_0)\rangle_T}{\langle w(x_n, x_{n-1}, \ldots, x_1, x_0)\rangle_T} \ . \end{aligned} \quad \text{(D.11)}$$

In a similar way, the average of every operator 0 which commutes with the Hamiltonian H (and so does with $G(H)$) is obtained by means of the following general formula:

$$\frac{\langle\psi_0|O|\psi_0\rangle}{\langle\psi_0|\psi_0\rangle} = \int \mathrm{d}x\, 0_L(x)\, \Pi_{eq}(x) \ .$$

where $0_L(x) = \frac{\langle\psi_T|O|x\rangle}{\langle\psi_T|x\rangle}$ is the local form of the observable. Because the presence of the weight $w(x, x_{n-1}, \ldots, x_1, x_0)$ in Eq. D.11, we shall call this kind of average *weighted averages*. They will be indicated by means of the symbol $\langle\cdots\rangle_w$.

The general case concerning operators O' which do not commute with the Hamiltonian is more complicated, but is not requested in the context of our work. We would like only to mention that the weighted average of O' is given by:

$$\langle 0'_L\rangle_w = \frac{\langle\psi_T|0'|\psi_0\rangle}{\langle\psi_T|\psi_0\rangle}$$

as one can check by using Eq. D.11. This is definitely different from the true ground state average $\langle\psi_0|O'|\psi_0\rangle$. Nevertheless more accurate estimates – having a good trial wavefunction – are known. In every practical case it is customary to employ the so-called *Ceperley's correction* [51, 64][pg. 48]:

$$\frac{\langle\psi_0|O'|\psi_0\rangle}{\langle\psi_0|\psi_0\rangle} \simeq 2\,\langle O'_L\rangle_w - \langle O'_L\rangle_T \ , \tag{D.12}$$

which is a second order approximation of the ground state average in the trial function accuracy (i.e. its error is $O\left(\|\psi_T - \psi_0\|\right)$. A even more precise way to estimate the ground state average is obtained by means of the so-called *forward-walker technique* [71, 72].

We shall see now how to employ the aforementioned continuous-time extended Markov chain to project the trial function ψ_T on the actual ground state ψ_0. The crucial point is the cumulative weight computation, which can be worked out in the following way: Suppose that a walker escapes from the configuration x after a time $t = n\,\Delta\,t$. The final weight will be:

$$b(x,t) = \lim_{n\to\infty}\left(\prod_{i=0}^{n-1} b(x_i)\right) = \lim_{n\to\infty}\left(1 - \Delta\,t\,(\epsilon_T(x) - \epsilon_T)\right)^n = e^{-t\,(\epsilon_L(x)-\epsilon_T)} \ . \tag{D.13}$$

As a consequence the weight cumulated during the continuous-time extended Markov chain $(x_n, t_n, \ldots, x_0, t_0)$ should be rewritten as (see Eq. D.10):

$$w(x_n,t_n) = w(x_n,t_n,\ldots,x_0,t_0) = \prod_{i=0}^{n-1} b(x_i,t_i) \ , \tag{D.14}$$

where we drop a part of the Markov chain indices for the sake of simplicity.

In summary, by means of a continuous-time Markov chain it is possible to implement the projection process defined in Eq. D.1 (i.e. the GFMC). The algorithm proceeds has follow: a continuous Markov chain is generated (for each walker) according to the transition operator defined in Eq. D.3. Then, the weight is updated and recorded as well as the local observable (e.g. $\epsilon_L(x)$) after every step, Because every Markov chain needs a finite amount of steeps (or time) before complete equilibration is reached (see Sec. 3.2), it is better to start data recording after a long enough equilibration time, t_{eq}. Moreover – because, in general, configurations are correlated – it is also better to record only data separated by an amount of Monte Carlo Steps (MC) corresponding, at least, to a correlation time t_{corr}. Both these times can be derived by warming-up simulations, for instance using the techniques that we explained in Sec. 3.2. We also recall that the time itself is a random variable and it must be incremented according to Eq. D.14.

At the end of the simulation an observable average, for instance the total energy average, is computed as follow:

$$\langle H\rangle = \langle\epsilon_L\rangle_w = \frac{\sum_i \sum_{j=1}^M \epsilon_L^{(j)}(x)\, w^{(j)}(x,t)}{\sum_i \sum_{j=1}^M w^{(j)}(x,t)} \ ,$$

where i is the (recorded) configuration counter and j the walker index.

Unfortunately such a scheme is plagued by a severe (average) weight divergence: $\langle\bar{w}(x,t)\rangle_T \to \infty$ as t increases. (where $\bar{w}(x,t) = \frac{\sum_{j=1}^M w^{(j)}(x,t)}{M}$ is the average of weight over the walker population).

We recall that the weight updating is a multiplicative process (see Eq. D.14) which depends on the local energy, $\epsilon_L(x) - \epsilon_T$ (see Eq. D.13). On the other hand, the local energy must have finite mean and finite variance (at the equilibrium), otherwise the whole Monte Carlo sampling is useless[48].

Without loss of generality, we can set: $\epsilon_T = \langle \epsilon_L \rangle_T$, and its variance equal to some (unknown) constant σ_T^2. It is also reasonable to suppose that the actual distribution of the local energy (at equilibrium) is Gaussian, especially whenever the trial function ψ_T is a good approximation of the actual ground state. Therefore the logarithm of the cumulative weight is also Gaussian, as it follows from the equation:

$$\ln w(x,t) = \sum_{i=0}^{n-1} \ln b(x_i, t_i) \simeq -\bar{t} \sum_{i=0}^{n-1} (\epsilon_l(x_i) - \epsilon_T) \ , \tag{D.15}$$

where $\bar{t}$ is the average time-increasing after a MCS (at the equilibrium). A good estimate of $\bar{t}$ is: $\bar{t} = \frac{1}{M} \sum_{j=1}^{M} \frac{1}{\lambda_x^{(j)}}$ (the average is obtained over the population). One can also see from Eq. D.15 that the means of $\ln w(x,t)$ is zero and its variance $\bar{t}^2 \, \sigma_T^2 \, (n-1)$. As a consequence the single weight $w(x,t)$ is distributed according to a *Lognormal* distribution [99][page 123ff], whose mean and variance are given by:

$$\begin{aligned} \langle w \rangle_T &= e^{\frac{\bar{t}^2 \, \sigma_T^2 \, (n-1)}{2}} \\ \langle w^2 \rangle_T - \langle w \rangle_T^2 &= \langle w \rangle_T^2 \left(e^{\bar{t}^2 \, \sigma_T^2 \, (n-1)} - 1 \right) \ . \end{aligned}$$

We stress that, if $\sigma_T > 0$, both these quantities diverge. A very accurate trial function or a large walker population can slower this divergence, but they can not avoid it, one can check by inspection, being the average weight, $\bar{w}$, mean and variance given by:

$$\begin{aligned} \langle \bar{w} \rangle_T &= e^{\frac{\bar{t}^2 \, \sigma_T^2 \, (n-1)}{2 \, M}} \\ \langle \bar{w}^2 \rangle_T - \langle \bar{w} \rangle_T^2 &= \langle \bar{w} \rangle_T^2 \left(e^{\bar{t}^2 \, \sigma_T^2 \, \frac{(n-1)}{M}} - 1 \right) \ . \end{aligned}$$

It is worth to note that, in the limit of infinite population, the mean of the average (over the population) weight is fixed and the correspondent variance is zero.

Because of such weights divergence, one may think that – for any practical purpose – the pure GFMC scheme which we explained so far is completely useless, However, a bunch of treatment have been devised and here we shall briefly sketch two of them.

First of all, one can apply the so-called *branching algorithm* [51, 64, 100]. In this case, instead of cumulate the weight as in eq. D.14, one stops the process after n_B MCS (or equivalently after a time $t_B = n_B \, \bar{t}$), and computes the observables averages. Then, every configuration is killed or multiplied according to the following equation:

$$N(x) = \lfloor u_3 + \sum_{i=1}^{M} \delta(x - x_i) \, w(x_i, t) \rfloor \ ,$$

where $N(x)$ is the number of walker in a given configuration x at the end of the branching, and $u_3 \in (0,1]$ is an equidistributed random number. In particular, $N(x) = 0$ means that all the walkers lying on configuration x have been deleted, while if $N(x) > 1$ they have been copied $N(x)$ times.

More precisely, the walker lying in configuration x is actually copied together all its characteristics (i.e. configuration, indeed, and also transition probabilities, observable averages, etc.). Finally all the weights are set to 1, as like at the beginning of the simulation. If the weight divergence did not happen during the previous n_B MCS (i.e. before the branching), such a process avoids to cumulate weight fluctuations in the further MCS, leading to enhanced stability. Nevertheless, the branching algorithm is an approximation of the true multiplicative process underlying the weight update. Therefore it is worth to note that – even if it reduces the observable fluctuation – it introduces an effective interaction among the walker, which was absent in the pure GFMC scheme. As a consequence, the observables averages taken just before the branching and immediately after may disagree:

$$\frac{\sum_{i+1}^{M} O(x_i)\, w(x_i,t)}{\sum_{i+1}^{M} w(x_i,t)} \neq \frac{\int \mathrm{d}x\, O(x)\, N(x)}{\int \mathrm{d}x\, N(x)} \,. \tag{D.16}$$

This bias is mainly due to the finite population M, since one can proves the convergence of the all averages as a power of $\frac{1}{M}$ [101]. As usual, by means of few initial simulations, one can find the optimal number of walkers that allows for fast and precise GFMC.

We are also interested in estimate the variance of the R.H.S. of Eq. D.16, since it is a measure of the bias we introduced through the branching algorithm. It is obtained by the sum of two contributions: The first one from the variance of the numerator (proportional to $\langle(\lfloor w(x)\rfloor - w(x))^2\rangle_T$) and the second one from the variance of the denominator (proportional to $\langle(\lfloor\sum w(x)\rfloor - \sum w(x))^2\rangle_T$). Anyway, we would like to stress again that, in the limit of infinity large population, the weight is constant and both the above contributions are zero. Therefore it is a good practice to use as many walkers as it is possible.

The denominator of Eq. D.16, R.H.S., is the average of the population after the branching:

$$M' = \int \mathrm{d}x\, N(x) \simeq \lfloor u_3 + \sum_{i=1}^{M} w(x_i,t)\rfloor \,,$$

which is no longer a constant. It is indeed customary to set the constant n_B (or t_B) in such a way to keep the population almost constant, bearing in mind that longer waiting-time causes larger weight fluctuations. We stress also that 1 minus the ratio of the deleted walker number (after the reconfiguration) over the population (before the reconfiguration) is the so-called *survival rate*. It is a measure of the efficiency of the branching process (i.e. if it is around 1, it means that the branching is useless, since almost nothing has changed. Moreover, it often means that the chosen value of the branching time, t_B, is too much small).

The second method is the so-called *stochastic reconfiguration* [71, 95, 102] It is an improvement of the branching algorithm devised to get rid of the bias due to the population fluctuations. In this case the number of walker lying in configuration x after every reconfiguration (i.e. branching) is given by:

$$N(x) = \lfloor u_3 + M\, \frac{\sum_{i=1}^{M} \delta(x - x_i)\, w(x_i,t_i)}{\sum_{i=1}^{M} w(x_i,t)} \rfloor \,,$$

where $u_3 \in (0,1]$ is the usual equidistributed random number.

It results that the walker number is effectively fixed, because

$$M' = \int \mathrm{d}x\, N(x) = \lfloor u_3 + M\rfloor = M \,,$$

and, since the population is constant, part of the variance of the average of observables taken after the reconfiguration (see Eq. D.16, R.H.S.) is zero.

For more details about stochastic reconfiguration, we refers to Ref. [71], where this topic is addressed in a very rigorous way.

D.2 A GFMC scheme

We make use of a multiwalker GFMC with stochastic reconfiguration scheme (see Sec. 5.4) in order to perform quantum annealing (QA) simulation of the two-dimensional Random Ising Model (see Chap. 5). Here we briefly sketch the employed algorithm. First of all, we employ the usual linear annealing schedule: $\Gamma = \Gamma_0 \left(1 - \frac{n}{\tau}\right)$, where n is the counter of the reconfiguration steps and τ the annealing time. We recall that a reconfiguration step is achieved after n_B Monte Carlo steps. We avoid to decrease the annealing parameter Γ at the end of every Monte Carlo step because it may cause weight instabilities (especially in the final part of the annealing). As a matter of convenience, in the following scheme we specify the difference between a true GFMC annealing and an equilibrium (i.e. with Γ fixed) GFMC algorithm (see Sec. 5.3).

Our GFMC scheme proceeds as follow:

Equilibration: We set the initial transverse field ($\Gamma = 2.5$ in the annealing case) and perform n_{eq} reconfiguration steps, every one consisting of n_B GFMC steps (see below). At the end of this part, all the internal counters and times are reset to zero, but configurations, transition probability, averages, and all the other quantities referring to the walker state are saved.

Main: It consists of τ reconfiguration steps, everyone made of n_B GFMC steps. In a equilibrium simulation Γ is kept constant, otherwise the annealing parameter is decreased according to the linear schedule: $\Gamma = \Gamma_0 \left(1 - \frac{n}{\tau}\right)$, where n is the reconfiguration steps counter.

At the early beginning of any reconfiguration step one has to compute the *average branching time*: $t_B = n_b \frac{1}{M} \sum_{i=1}^{M} \frac{1}{\lambda_x^{(i)}}$, and one also must set to zero the *internal time* $t^{(i)}$ of every walker. Then an independent series of GFMC steps for each walker is performed. This is consistent to the theory, which says that the walkers are independent till a reconfiguration (or a branching) is attained (see Sec.D.1).

Every GFMC steps consists of:

1. First of all, the internal time of the i-th (we usually start from $i = 1$) walker is increased by: $t_{incr} = -\frac{\ln u_1}{\lambda_{exc}^{(i)}(x)}$, where $u_1 \in (0, 1]$ is an equidistributed random number (see Eq. D.5).
2. Then a new configuration x_i' is chosen according to the method described in Eq. D.6 (this operation requires another equidistributed random number).
3. Finally, the weight, the observables, and the transition probability are updated according to the new configuration reached by the walker.

If at the end of this step the internal time is still lesser then the branching time: $t^{(i)} < t_b$, another GFMC for *the same walker* is performed. Otherwise a GFMC step for the $(i+1)$-th walker can be started and so long so forth, till all the walkers are updated.

At the end of every reconfiguration step (i.e. after an average of n_B GFMC for every walker) the global time is increased by t_B, the observables means (over the walker population), $\bar{0}$, are computed according to the general formula (see the L.H.S. of Eq. D.16), and finally they are recorded. Also the average weight $\bar{w} = \frac{\sum_{j=1}^{M} w^{(j)}(x,t)}{M}$ must be saved at the end of every step. Then the stochastic reconfiguration procedure is performed (see Sec. D.1) and the walkers (their configuration, transition probability, etc.) are copied or deleted as a consequence. Only after this point all the weight set to 1 and another reconfiguration step may start.

That is all for an annealing simulation. The averages we saved can not be weighted again, because they are obtained at different transverse fields, Γ. As a consequence – in order to have a suitable statistics – one must repeat the whole annealing algorithm several times (starting from a different initial condition or changing the random number generator seed), and eventually one has to take the final averages and variances over such repetitions. For this reason an annealing simulation is really time consuming.

On the other hand – for an equilibrium simulation – many *equivalent* observable averages have been recorded, since the external condition (i.e. the transverse magnetic field Γ) has been fixed. Therefore one can increase the statistics simply making very long simulations, and collect more configurations (provided that they are uncorrelated, see previous appendix). The final *weighted* averages are then obtained by means of the observables means $\bar{O}$ and the average weight $\bar{w}$ recorded at the end of every reconfiguration step (we remember that such averages have been taken over the walker population). The final weighted averages and variances can be worked out by means of the following formulas:

$$\langle \bar{O} \rangle = \frac{\sum_i \bar{O}_i \, \bar{w}_i}{\sum_i \bar{w}_i}, \tag{D.17}$$

$$\langle \bar{O}^2 \rangle - \langle \bar{O} \rangle^2 = \frac{\sum_i \bar{O}_i^2 \, \bar{w}_i}{\sum_i \bar{w}_i} - \langle \bar{O} \rangle^2 \, . \tag{D.18}$$

It is also clear that is much more convenient to perform a long simulation instead of rerun from scratch many time, as like for the annealing case. In fact one can avoid this way many useless initial equilibration steps.

Bibliography

[1] Jr. S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671, 1983. (document), 3

[2] A.B. Finnila, M.A. Gomez, C. Sebenik, C. Stenson, and J.D. Doll. Quantum annealing: A new method for minimizing multidimensional functions. *Chem. Phys. Lett.*, 219:343, 1994. (document), 3

[3] T. Kadowaki and H. Nishimori. Quantum annealing in the transverse Ising model. *Phys. Rev. E*, 58:5355, 1998. (document), 1.2, 2.2.1, 32

[4] J. Brooke, D. Bitko, T.F. Rosenbaum, and G. Aeppli. Quantum annealing of a disordered magnet. *Science*, 284:779, 1999. (document), 1.1, 33

[5] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda. A quantum adiabatic evolution algorithm applied to random instances of an NP-Complete problem. *Science*, 292:472, 2001. (document)

[6] A. Das, B. Chakrabarti, and R.B. Stinhcombe. Quantum annealing in a kinetically constrained system. *Phys. Rev. E*, 72:026701, 2005. (document)

[7] P. Amara, D. Hsu, and J.E. Straub. Global energy minimum searches using an approximate solution of the imaginary time Schrödinger equation. *J. Phys. Chem.*, 97:6715, 1993. (document)

[8] Y.H. Lee and B.J. Berne. Global optimization: Quantum thermal annealing with path integral monte carlo. *J. Phys. Chem. A*, 104:86, 2000. (document), 3

[9] P. Liu and B.J. Berne. Quantum path minimization: An efficient method for global optimization. *j. Chem. Phys.*, 118:2999, 2003. (document), 3, 4

[10] G.E Santoro, R. Martoňák, E. Tosatti, and R. Car. Theory of quantum annealing of an Ising spin glass. *SCIENCE*, 295:2427, 2002. (document), 1.1, 1.1, 1.1, 3, 4, 32, 32, 5.3, 5.4, 49, 5.5

[11] R. Martoňák, G.E. Santoro, and E. Tosatti. Quantum annealing by the path-integral Monte Carlo method: The two-dimensional random Ising model. *Phys. Rev. B*, 66:094203, 2002. (document), 1, 1.1, 1.1, 4, 32, 32, 5.5

[12] Y.H. Lee and B.J. Berne. Quantum thermal annealing with renormalization: Application to a frustrated model protein. *J. Phys. Chem. A*, 105:2001, 459. (document), 3

[13] R. Martoňák, G.E. Santoro, and E. Tosatti. Quantum annealing of the traveling-salesman problem. *Phys. Rev. E*, 70:057701, 2004. (document), 1.2, 1.2, 1.2, 1.3, 1.2, 1, 3, 4

[14] P.F. Stadler and W. Schnabl. The landscape of the traveling salesman problem. *Phys. Lett. A*, 161:337, 1992. (document)

[15] O. Dubois, R. Monasson, B. Selman, and R. Zecchina (Editors). Special issue on NP-hardness and phase transitions. *Theor. Comp. Sci.*, 265(1-2), 2001. (document), 1, 2

[16] D. Battaglia, G.E. Santoro, and E. Tosatti. Optimization by quantum annealing: Lessons from hard satisfiability problems. *Phys. Rev. E*, 71:066707, 2005. (document), 1.3, 1.4, 2, 4, 50

[17] L. Stella, G.E. Santoro, and E. Tosatti. Optimization by quantum annealing: Lessons from simple cases. *Phys. Rev. B*, 72:014303, 2005. (document)

[18] C.H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Dover, 1998. 1

[19] H. Nishimori. *Statistical Physics of Spin Glasses and Information Processing*. Oxford University Press, 2001. 1

[20] J.J. Hopfield and D.W. Tank. Computing with neural circuits: A model. *Science*, 233:625, 1986. 1, 1.2

[21] M. Mézard and G. Parisi. Mean-field equations for the matching and travelling salesman problems. *Europhys. Lett.*, 2:913, 1986. 1, 1.2

[22] A.G. Percus and O.C. Martin. The stochastic traveling salesman problem: Finite size scaling and the cavity prediction. *Jour. Stat. Phys.*, 94:739, 1999. 1, 1.2

[23] D.S. Dean, D. Lancaster, and S.N. Majumdar. The statistical mechanics of travelling salesman type problems. *J. Stat. Mech.*, page L01001, January 2005. 1, 1.2

[24] S. Mertens. Phase transition in the number partitioning problem. *Phys. Rev. Lett*, 81:4281, 1988. 1

[25] M. Mézard, G. Parisi, and R. Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297:812, 2002. 1, 2

[26] M. Mézard and R. Zecchina. Random k-satisfiability problem: From an analytic solution to an efficient algorithm. *Phys. Rev. E*, 66:56126, 2002. 1, 2

[27] M. Weigt and A.K. Hartmann. Number of guards needed by a museum: A phase transition in vertex covering of random graphs. *Phys. Rev. Lett*, 84:2000, 6118. 1

[28] A. Braunstein, R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina. Polynomial iterative algorithms for coloring and analyzing random graphs. *Phys. Rev. E*, 68:036702, 2003. 1

[29] G. Parisi, M. Mézard, and M.Virasoro. *Spin Glass Theory and Beyond*. World Scientific, 1987. 1, 45

[30] C.H. Papadimitriou. *Computational complexity*. Addison Wesley, 1993. 1

[31] M.R. Garey and D.S. Johnson. *Computers and Intractability*. Freeman, 1979. 1

[32] D. Mitchell, B. Selman, and H. Levesque. Hard and easy distribution of sat problems. In *Proc. of the Tenth National Conference on Artificial Intelligence*, page 459. AAAI-Press, 1992. 1

[33] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina. Two solutions to diluted p-spin models and xorsat problems. *J. Stat. Phys*, 111:505, 2003. 1

[34] F. Barahona. On the computational complexity of Ising spin glass models. *J. Phys. A*, 15:3241, 1982. 1.1

[35] C. De Simone, M. Diehl, M. Jünger, P. Mutzel, G. Reinelt, and G. Rinaldi. Exact ground states of Ising spin glasses: New experimental results with a branch-and-cut algorithm. *J. Stat. Phys.*, 80:487, 1995. 1.1, 2

[36] D.S. Johnson and L.A. McGeoch. The travelling salesman problem: A case study. In *Local Search in Combinatorial Optimization*, page 215. Princeton University Press, 2003. 1.2, 1

[37] P.F. Stadler. Towards a theory of landscapes. In R. Lopez-Pena, R. Capovilla, R. Garca-Pelayo, H. Waelbroeck, and F. Zertuche, editors, *Complex Systems and Binary Networks*, page 77. Springer Verlag, 1995. 1.2

[38] D. Battaglia, M. Kolář, and R. Zecchina. Minimizing energy below the glass thresholds. *Phys. Rev. E*, 70:036107, 2004. 2

[39] A. Montanari, G. Parisi, and F. Ricci-Tersenghi. Instability of one-step replica-symmetry-broken phase in satisfiability problems. *J. Phys. A*, 37:2073, 2004. 2

[40] Z. Michalewicz. *Genetic algorithms + data structures = evolution programs*. Springer-Verlag, third edition, 1996. 2

[41] N.G. van Kampen. *Stochastic processes in physics and chemistry*. North-Holland, revised and enlarged edition edition, 1992. 3, 6, 3.1, 8, 3.4, 14, 10

[42] S. Shinomoto and Y. Kabashima. Finite time scaling of energy in simulated annealing. *J. Phys. A*, 24:L141, 1991. 2.1.1, 2.3, 2.3, 2.3, A.1.1

[43] Y. Kabashima and S. Shinomoto. Asymptotic dependence of the residual energy on annealing time. *J. Phys. Soc. Japan*, 60:3993, 1991. 2.1.1, 2.3, 2.3, 2.3, A.1.1

[44] D.A. Huse and D.S. Fisher. Residual energies after slow cooling of disordered system. *Phys. Rev. Lett.*, 57:2203, 1986. 5, 7, 2

[45] L.D. Landau. A theory of energy transfer II. *Phys. Z. Sowjetunion*, 2:46, 1932. 6

[46] C. Zener. Non-adiabatic crossing of energy levels,. *Proc. Royal Soc. A*, 137:696, 1932. 6

[47] P.A. Lee and T.V. Ramakrishnan. Disordered electronic systems. *Rev. Mod. Phys.*, 57:287, 1985. 2.4

[48] D.P. Landau and K. Binder. *A guide to Monte Carlo simulations in statistical physics*. Cambridge University Press, 2000. 3, D.1

[49] F.R. Gantmakher. *The theory of matrix*. Chelsea Pub. Co., 1960. 3.1, 8, 8, A.3

[50] N. Metropolis, A. W. Rosenbloth, M. N. Rosenbloth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087, 1953. 3.2, 3.3

[51] W.M.C. Foulkes, L. Mitas, R.J. Needs, and G. Rajagopal. Quantum Monte Carlo simulation of solids. *Rev. Mod. Phys.*, 73:33, 2001. 3.2, 50, D.1, D.1, D.1, D.1

[52] H. Flyvbjerg and H.G. Petersen. Error estimates on averages of correlated data. *J. Chem. Phys.*, 91:461, 1989. 12, 5, 13, 2

[53] M.E.J. Newman and G.T. Barkema. *Monte Carlo methods in statistical physics*. Clarendon Press, first edition, 1999. 12, C.4

[54] H. Risken. *The Fokker-Planck equation: methods of solution and applications*. Springer-Verlag, second edition, 1989. 3.4

[55] S. Szu and R. Hartley. Fast simulating annealing. *Phys. Lett. A*, 122:157, 1987. 16

[56] Y. Xiang and X.G. Gong. Efficiency of the generalized simulated annealing. *Phys. Rev. E*, 62:4473, 2000. 16

[57] C. Tsallis. Possible generalization of boltzmann-gibbs statistics. *J. Stat. Phys.*, 52:479, 1988. 16

[58] R.P. Feynman. *Statistical mechanics: a set of lectures*. Benjamin, 1972. 4.1

[59] D.M. Ceperley. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.*, 67:279, 1995. 3, 4, C.3, 59, C.5, C.6, 60

[60] Staging: A sampling technique for the Monte Carlo evaluation of path integrals. M. sprik and m.l. klein and d. chandler. *Phys. Rev. B*, 31:4234, 1985. 4

[61] D.S. Fisher and D.A. Huse. Ordered phase of short-range Ising spin-glasses. *Phys. Rev. Lett.*, 56:1601, 1986. 2

[62] J. Brooke, T.F. Rosenbaum, and G. Aeppli. Tunable quantum tunnelling of magnetic domain walls. *Nature*, 413:610, 2001. 33

[63] S. Sorella, G.E. Santoro, and F. Becca. Sissa lecture notes on numerical methods for strongly correlated electrons, 2004. 5.1, 38, 40

[64] N. Trivedi and D.M. Ceperley. Ground state correlations of quantum antiferromagnets: A Green-function Monte Carlo study. *Phys. Rev. B*, 41:4552, 1990. 40, D.1, D.1

[65] S.Sorella and L. Capriotti. Green function monte carlo with stochastic reconfiguration: An effective remedy for the sign problem. *Phys. Rev. B*, 61:2599, 2000. 40

[66] K. Binder. Introduction: Theory and "technical" aspects of Monte Carlo simulations. In Second, editor, *Monte Carlo methods in statistical physics*, page 1. Springer-Verlag, 1986. 40, 40

[67] H. Rieger and A.P. Young. Zero-temperature quantum phase transition of a two dimensional Ising spin glass. *Phys. Rev. Lett.*, 72:4141, 1994. 10

[68] K. Binder and A.P. Young. Spin glasses: Experimental fact, theoretical concepts, and open questions. *Rev. Mod. Phys.*, 58:801, 1986. 41

[69] T. Castellani and A. Cavagna. Spin-glass theory for pedestrians. *J. Stat. Mech*, page P05012, May 2005. 45

[70] S. Sorella. Wave function optimization in Variational Monte Carlo. *Phys. Rev. B*, 71:241103(R), 2005. 19, D.1, D.1

[71] M. Calandra Buonaura and S. Sorella. Numerical study of the two-dimensional Heisenberg model using a Green function Monte Carlo technique with a fixed number of walkers. *Phys. Rev. B*, 57:11446, 1998. 50, D.1, D.1

[72] B.L. Hammond, W.A. Lester, and P.J. Reynolds. *Monte Carlo methods in ab initio quantum chemistry*. World Scientific, 1994. 50, D.1, D.1, D.1

[73] G. Parisi. *Statistical field theory*. Addison-Wesley, 1988. A.2, A.2, A.2

[74] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical recipes in C: the art of scientific computing*. Cambridge, University Press, second edition, 1992. A.3, A.3, C.4, 2

[75] S. Tănase-Nicola and J. Kurchan. Metastable states, transitions, basins and borders at finite temperatures. *J. Stat. Phys.*, 116:1201, 2004. B.2, 57

[76] S. Tănase-Nicola and J. Kurchan. Topological methods for searching barriers and reaction paths. *Phys.Rev.Lett.*, 91:188302, 2003. B.2, 57

[77] J.W. Negele and H. Orland. *Quantum many-particle systems*. Addison-Wesley, 1988. B.3

[78] H. Kleinert. *Path integrals in quantum mechanics, statistics, polymer physics and financial markets Hagen*. World Scientific, third edition, 2004. 54

[79] M. Le Bellac. *Quantum and statistical field theory*. Clarendon Press, 1991. C.1

[80] T. Kashiwa, Y. Ohnuki, and M. Suzuki. *Path integral methods*. Oxford Clarendon Press, 1997. C.2

[81] M. Suzuki. General theory of higher-order decomposition of exponential operators and symplectic integrators. *Phys. Lett. A*, 165:387, 1992. C.2

[82] M. Suzuki. General decomposition theory of exponential operators. In M. Suzuki, editor, *Quantum Monte Carlo methods in condensed matter physics*. World Scientific, 1993. C.2

[83] N. hatano and M. Suzuki. Finding exponential product formulas of higher order. `math-ph/0506007`. C.2

[84] M. Takahashi and M. Imada. Monte Carlo calculation of quantum system. II. Higher order correction. *J. Phys. Jap.*, 53:3765, 1984. C.2

[85] X.-P. Li and J.Q. Broughton. High-order correction to the Trotter expansion for use in computer simulation. *J. Chem. Phys.*, 86:5094, 1987. C.2, C.2

[86] L.D. Landau, E.M. Lifshits, and L.P. Pitaevskii. *Statistical physics*, volume 5 of *Course of theoretical physics*. Pergamon Press, third edition, 1980. C.2

[87] S. Jang, S. Jang, and G.A. Voth. Applications of higher order composite factorization scheme in imaginary time path integral simulations. *J. Chem. Phys.*, 115:7832, 2001. C.2

[88] T.M. Yamamoto. Path integral virial estimator based on the scaling of fluctuation coordinates: Application to quantum clusters with fourth-order propagators. `cond-mat/0505109`. C.2

[89] L. Brualla, K. Sakkos, J. Boronat, and J. Casulleras. Higher order and infinite trotter-number extrapolations in path integral monte carlo. *J. Chem. Phys.*, 121:636, 2004. C.2

[90] M.F. Herman, E.J. Bruskin, and B.J. Berne. On path integral Monte Carlo simulations. *J. Chem. Phys.*, 76:5150, 1982. C.3

[91] L.D. Landau and E.M. Lifshits. *Mechanics*, volume 1 of *Course of theoretical physics*. Pergamon Press, third edition, 1976. 59

[92] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: A fractional dynamics approach. *Phys. Rep.*, 339:1, 2000. C.4, C.12

[93] J. Kafler, A. Blumen, and M.F. Shlesinger. Stochastic pathway to anomalous diffusion. *Phys. Rev. A*, 35:3081, 1987. C.4, C.12

[94] S. Eidelman et al. The review of particle physics. *Phys. Lett. B*, 592:1, 2004. C.12, C.12

[95] R. Assaraf, M. Caffarel, and A. Khelif. Diffusion monte carlo with a fixed number of walker. *Phys. Rev. B*, 61:4566, 2000. D.1, D.1

[96] X. Lin, H. Zhang, and A. Rappe. Optimization of quantum Monte Carlo wave functions using analytical energy derivatives. *J. Chem. Phys.*, 112:2650, 2000. D.1, D.1

[97] K. A. Fichthorn and W.H. Weinberg. Theoretical foundations of dynamical Monte Carlo simulations. *J. Chem. Phys.*, 95:1090, 1991. D.1

[98] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340, 1977. 1, 2

[99] J.F. Kenney and E.S. Keeping. *Mathematics of Statistics*, volume 2. Van Nostrand, second edition, 1951. D.1

[100] P.J. Reynolds, D.M. Ceperley, B.J. Adler, and Jr. W.A. Lester. Fixed-node quantum Monte Carlo for molecules. *J. Chem. Phys.*, 77:5593, 1982. D.1

[101] M. Beccaria. Green function simulation of Hamiltonian lattice model with stochastic reconfiguration. *Eur. Phys. J. C.*, 13:357, 2000. D.1

[102] J.H. Hetherington. Observations on the statistical iteration of matrices. *Phys. Rev. A*, 30:2713, 1984. D.1

Acknowledgments

I want to thank Giuseppe Santoro which supervised my work during the four years that I spent in SISSA. I also express my gratitude to Erio Tosatti for the support and the important suggestions that he gave me from the early beginning. Many other people provided valuable help in several different ways: Llorenc Brualla explained me all I know about PIMC and Saverio Moroni discussed with me about GFMC. I previously reported part of this work at the first Quantum Annealing Conference, held in Kolkata in March. I want to mention the kind hospitality of the organizers: Prof. Bikas Chakrabarti and Dr. Arnab Das.

In Trieste I learnt a lot about Physics and, even more important, a lot about life. For the first time I lived alone, far from my family, and I began to take care of my little business, and of my clothes. Nevertheless I was very lucky, since I found a large amount of new and real friends. They helped me in the bad time (when I was in the hospital) and in the good time (thanks to them, almost every day!). I want to thank: Fabio and Simona (I'm writing on the Fabio's laptop, indeed), Claudio "W." and Michele "the colonel" (we were always in the same office), Simone (he hosted me in Rome for the first, celebrated, "Rome meets SISSA"), Lorenzo (it is not worth to refer anything about our bad jokes...) and Katrin, Tania (poor Tania, she tried to teach me Russian!) and Evgeny, Oleg, Laura (since she laughs whenever I say something) and Pushpa, Giacomo (we shared also the same flat for a while), Carlo (he makes me proud of my Venetian origin) and Valentina, Vittoria and Hamed (especially for its sense of humor), Demian (he bore me in India and he introduced me in the spin-glass theory: definitely two very difficult tasks!) and Osvaldo, Federica (and her mother, too), Davide and Ugo (the best in SISSA!), Sara T. and Francesca (they are not the same person!), Stefano and Manuela (my new office-mates), Pasquale, Željka, Francesco, Bruno, and Iryna (among the astrophysics), Lucattilio, Fabio P., Chiara P., and Massimo (among the mathematicians), Maria Sole and Lorena (in particular,

she gave me a lift many times), Sara B. (she hosted many of us in her flat), and many others (I'm sorry, this list is becoming to much long!).

Many thanks also to all the Adriatico staff (Sandro, Patrizia, Lucia, and Tiziano) for the delicious food provided. Probably the lunch break was the best time that I spent in Miramare campus.

Finally I want to recall two basis facts: Without the help of my family (especially my mother) nothing of that would by possible. I like physics and I spent a lot of energy to be a physicist, but I finally understood that it is not the most important thing in my life. Chiara knows why.